

Syst Synth Biol (2007) 1:59–87
DOI 10.1007/s11693-007-9005-9

REVIEW

Monotone and near-monotone biochemical networks

Eduardo D. Sontag

Received: 17 December 2006 / Accepted: 19 March 2007 / Published online: 4 July 2007
© Springer Science+Business Media B.V. 2007

Abstract Monotone subsystems have appealing properties as components of larger networks, since they exhibit robust dynamical stability and predictability of responses to perturbations. This suggests that natural biological systems may have evolved to be, if not monotone, at least close to monotone in the sense of being decomposable into a “small” number of monotone components. In addition, recent research has shown that much insight can be attained from decomposing networks into monotone subsystems and the analysis of the resulting interconnections using tools from control theory. This paper provides an expository introduction to monotone systems and their interconnections, describing the basic concepts and some of the main mathematical results in a largely informal fashion.

Keywords Biochemical networks · Dynamical systems · Monotone systems

Introduction

The field of systems molecular biology is largely concerned with the study of biochemical networks consisting of proteins, RNA, DNA, metabolites, and other molecules. These networks participate in control and signaling in development, regulation, and metabolism, by processing environmental signals, sequencing internal events such as gene expression, and producing appropriate cellular responses. It is of great interest to be able to infer dynamical

properties of a biochemical network through the analysis of well-characterized subsystems and their interconnections. This paper discusses recent work which makes use of both *topology* (graph structure) and *sign* information in order to deduce such properties.

It is broadly appreciated that behavior is critically dependent on network topology as well as on the signs (activating or inhibiting) of the underlying feedforward and feedback interconnections (Novic and Weiner 1957; Monod and Jacob 1961; Lewis et al. 1977; Segel 1984; DeAngelis et al. 1986; Thomas and D’ari 1990; Goldbeter 1996; Keener and Sneyd 1998; Murray 2002; Milo et al. 2002; Edelstein-Keshet 2005). For example, Fig. 1a–c shows the three possible types of feedback loops that involve two interacting chemicals. A mutual activation configuration is shown in Fig. 1a: a positive change in *A* results in a positive change in *B*, and vice-versa. Configurations like these are associated to signal amplification and production of switch-like biochemical responses. A mutual inhibition configuration is shown in Fig. 1b: a positive change in *A* results in repression of *B*, and repression of *B* in turn enhances *A*. Such configurations allow systems to exhibit multiple discrete, alternative stable steady-states, thus providing a mechanism for memory. Both (a) and (b) are examples of positive-feedback systems (Ptashne 1992; Plahte et al. 1995; Cinquin and Demongeot 2002; Gouze 1998; Thomas and Kaufman 2001; Remy et al. 2003; Angeli and Sontag 2004a; Angeli et al. 2004a). On the other hand, activation-inhibition configurations like in Fig. 1c are necessary for the generation of periodic behaviors such as circadian rhythms or cell cycle oscillations, by themselves or in combination with multi-stable positive-feedback subsystems, as well as for adaptation, disturbance rejection, and tight regulation (homeostasis) of physiological variables

Supported in part by NSF Grants DMS-0504557 and DMS-0614371.

E. D. Sontag (✉)
Rutgers University, New Brunswick, NJ, USA
e-mail: sontag@math.rutgers.edu

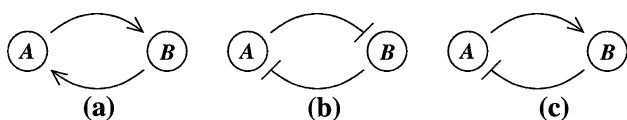


Fig. 1 (a) Mutual activation. (b) Mutual inhibition. (c) Activation-inhibition

(Goldbeter 1996; Murray 2002; Edelstein-Keshet 2005; Rapp 1975; Hastings et al. 1977; Tyson and Othmer 1978; Thomas 1981; Sontag 1998; Kholodenko 2000; Sha et al. 2003; Pomerening et al. 2003; Angeli and Sontag 2004b). Compared to positive-feedback systems, negative-feedback systems are not “consistent,” in a sense to be made precise below but roughly meaning that different paths between any two nodes should reinforce, rather than contradict, each other. For (c), a positive change in A will be resisted by the system through the feedback loop. Consistency, or lack thereof, also plays a role in the behavior of graphs without feedback; for example Milo et al. (2002), Mangan and Alon (2003), Mangan et al. (2003) deal with the different signal processing capabilities of consistent (“coherent”) compared to inconsistent feedforward motifs.

A key role in the work to be discussed here will be played by consistent systems and subsystems. We will discuss the following points:

- Interesting and nontrivial conclusions can be drawn from (signed) network structure alone. This structure is associated to purely stoichiometric information about the system and ignores fluxes. Consistency, or close to consistency, is an important property in this regard.
- Interpreted as dynamical systems, consistent networks define *monotone systems*, which have highly predictable and ordered behavior.
- It is often useful to analyze larger systems by viewing them as interconnections of a small number of monotone subsystems. This allows one to obtain precise bifurcation diagrams without appeal to explicit knowledge of fluxes or of kinetic constants and other parameters, using merely “input/output characteristics” (steady-state responses or DC gains). The procedure may be viewed as a “model reduction” approach in which monotone subsystems are viewed as essentially one-dimensional objects.
- The possibility of performing a decomposition into a small number of monotone components is closely tied to the question of how “near” a system is to being monotone.
- We argue that systems that are “near monotone” are biologically more desirable than systems that are far from being monotone.
- There are indications that biological networks may be much closer to being monotone than random networks that have the same numbers of vertices and of positive and negative edges.

The need for robust structures and robust analysis tools

In contrast to many areas of applied mathematics and engineering, the study of dynamics in cell biology should take into account the often huge degree of uncertainty inherent in models of cellular biochemical networks, which arises from environmental fluctuations or from variability among cells of the same type. From a mathematical analysis perspective, this uncertainty translates into the difficulty of measuring the relevant model parameters such as kinetic constants or cooperativity indices, and hence the impossibility of obtaining a precise model.

This means that it is important to develop tools that are “robust” in the sense of being able to lead to useful conclusions from information regarding the *qualitative* features of the network, and, if possible, not upon the precise values of parameters or even the forms of reactions. This goal is hard to attain, since dynamical behavior may be subject to phase transitions (bifurcations) which critically depend on parameter values. Nevertheless, and perhaps surprisingly, there have been many successes in finding rich classes of chemical network structures for which such robust analysis is indeed possible. One approach is that of graph-theoretic ideas associated to *complex balancing and deficiency theory*, pioneered by Clarke (1980), Horn and Jackson (1972, 1974), Feinberg and Horn (1974), and Feinberg (1987, 1995). Another approach, pioneered by Hirsch and Smith, see Smith (1995), Hirsch and Smith (2005), relies upon the theory of *monotone systems*, and has a similar goal of drawing conclusions about dynamical behavior based only upon structure. This direction has been enriched substantially by the introduction of *monotone systems with inputs and outputs*: as standard in control theory (Sontag 1998), one extends the notion of monotone system so as to incorporate input and output channels (Angeli and Sontag 2003). Once inputs and outputs are introduced, one can study interconnections of systems (Fig. 2), and ask what special properties hold if the subsystems are monotone (Angeli et al. 2004a; Angeli and Sontag 2003; de Leenheer et al. 2007).

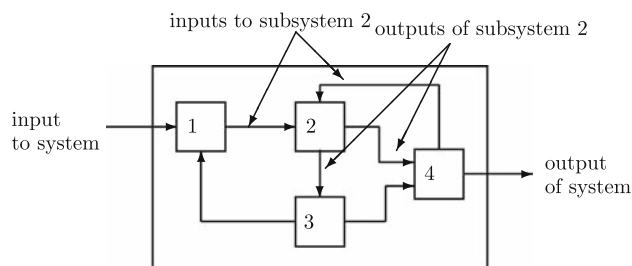


Fig. 2 A system composed of four subsystems

Consistent graphs, monotone systems, and near-monotonicity

We now introduce the basic notions of monotonicity and consistency. The present section deals exclusively with graph-theoretic information, which is derived from stoichiometric constraints. Complementary to this analysis, bifurcation phenomena can be sometimes analyzed using a combination of these graphical techniques together with information on steady-state gains; that subject is discussed in section “I/O monotone systems.” In order to preserve readability, the discussion in this section is informal, and not all mathematical technicalities are explained; references are given that will allow the reader to fill-in the missing details, and also section “I/O monotone systems” has more rigorous mathematical statements, presented in the more general context of systems with external inputs and outputs.

The systems considered here are described by the evolution of states, which are time-dependent vectors $x(t) = (x_1(t), \dots, x_n(t))$ whose components x_i represent concentrations of chemical species such as proteins, mRNA, or metabolites. In autonomous differential equation (“continuous-time”) models, one specifies the rate of change of each variable, at any given time, as a function of the concentrations of all the variables at that time:

$$\begin{aligned}\frac{dx_1}{dt}(t) &= f_1(x_1(t), x_2(t), \dots, x_n(t)) \\ \frac{dx_2}{dt}(t) &= f_2(x_1(t), x_2(t), \dots, x_n(t)) \\ &\vdots \\ \frac{dx_n}{dt}(t) &= f_n(x_1(t), x_2(t), \dots, x_n(t)),\end{aligned}$$

or just $dx/dt = f(x)$, where f is the vector function with components f_i . We assume that the coordinates x_i of the state of the system can be arbitrary non-negative numbers. (Constraints among variables can be imposed as well, but several aspects of the theory are more subtle in that case.) Often, one starts from a differential equation system written in the following form:

$$\frac{dx}{dt}(t) = \Gamma R(x),$$

where $R(x)$ is a q -dimensional vector of reactions and Γ is an $n \times q$ matrix, called the *stoichiometry matrix*, and either one studies this system directly, or one studies a smaller set of differential equations $dx/dt = f(x)$ obtained by eliminating variables through the use of conserved stoichiometric quantities.

We will mostly discuss differential equation models, but will also make remarks concerning difference equation (“discrete time”) models. The dynamics of these are described by rules that specify the state at some future time

$t = t_{k+1}$ as a function of the state of the system at the present time t_k . Thus, the i th coordinate evolves according to an update rule:

$$x_i(t_{k+1}) = f_i(x_1(t_k), x_2(t_k), \dots, x_n(t_k))$$

instead of being described by a differential equation. Usually, $t_k = k\Delta$ where Δ is a uniform inter-sample time. One may associate a difference equation to any given differential equation, through the rule that the vector $x(t_{k+1})$ should equal the solution of the differential equation when starting at state $x(t_k)$. However, not every difference equation arises from a differential equation in this manner. Difference equations may be more natural when studying processes in which measurements are made at discrete times, or they might provide a macroscopic model of an underlying stochastic process taking place at a faster time scale.

One may also study more complicated descriptions of dynamics than those given by ordinary differential and difference equations; many of the results that we discuss here have close analogs that apply to more general classes of (deterministic) dynamical systems, including reaction–diffusion partial differential equations, which are used for space-dependent problems with slow diffusion and no mixing, delay-differential systems, which help model delays due to transport and other cellular phenomena in which concentrations of one species only affect others after a time interval, and integro-differential equations (Smith 1995; Hirsch and Smith 2005; Sontag 2004; Enciso et al. 2006). In a different direction, one may consider systems with external inputs and outputs (Angeli and Sontag 2003).

The graph associated to a system

There are at least two types of graphs that can be naturally associated to a given biochemical network. One type, sometimes called the *species-reaction graph*, is a bipartite graph with nodes for reactions (fluxes) and species, which leads to useful analysis techniques based on Petri net theory and graph theory (Feinberg 1991; Reddy et al. 1993; Zevedei-Oancea and Schuster 2003; Craciun and Feinberg 2005; Craciun and Feinberg 2006; Angeli and Sontag 2007; Angeli et al. 2006, 2007). We will not discuss species-reaction graphs here. A second type of graph, which we will discuss, is the *species graph* G . It has n nodes (or “vertices”), which we denote by v_1, \dots, v_n , one node for each species. No edge is drawn from node v_j to node v_i if the partial derivative $\partial f_i / \partial x_j(x)$ vanishes identically, meaning that there is no direct effect of the j th species upon the i th species. If this derivative is not identically zero, then there are three possibilities: (1) it is ≥ 0 for all x , (2) it is ≤ 0 for all x , or (3) it changes sign depending on the particular

entries of the concentration vector x . In the first case (activation), we draw an edge labeled $+$, $+1$, or just an arrow \rightarrow . In the second case (repression or inhibition), we draw an edge labeled $-$, -1 , or use the symbol \vdash . In the third case, when the sign is ambiguous, we draw both an activating and an inhibiting edge from node v_j to node v_i . The graph G is an example of a *signed graph* (Zaslavsky 1998), meaning that its edges are labeled by signs.

For continuous-time systems, no self-edges (edges from a node v_i to itself) are included in the graph G , whatever the sign of the diagonal entry $\partial f_i / \partial x_i$ of the Jacobian. For discrete-time systems, on the other hand, self-edges are included (we later discuss the reason for these different definitions for differential and difference equations).

When working with graphs, it is more convenient (though not strictly necessary) to consider only graphs G that have no multiple edges from one node to another (third case above). One may always assume that G has this property, by means of the following trick: whenever there are two edges, we replace one of them by an indirect link involving a new node; see Fig. 3. Introducing such additional nodes if required, we will suppose from now on that no multiple edges exist.

Although adding new edges as explained above is a purely formal construction with graphs, it may be explained biologically as follows. Often, ambiguous signs in Jacobians reflect heterogeneous mechanisms. For example, take the case where protein A enhances the transcription rate of gene B if present at high concentrations, but represses B if its concentration is lower than some threshold. Further study of the chemical mechanism might well reveal the existence of, for example, a homodimer that is responsible for this ambiguous effect. Mathematically, the rate of transcription of B might be given algebraically by the formula $k_2 a^2 - k_1 a$, where a denotes the concentration of A. Introducing a new species C to represent the homodimer, we may rewrite this rate as $k_2 c - k_1 a$, where c is the concentration of C, plus an new equation like $dc/dt = k_3 a^2 - k_4 c$ representing the formation of the dimer and its degradation. This is exactly the situation in Fig. 3.

Spin assignments and consistency

A *spin assignment* Σ for the graph G is an assignment, to each node v_i , of a number σ_i equal to “+” or “−1” (a “spin,” to borrow from statistical mechanics terminology).



Fig. 3 Replacing direct inconsistent effects by adding a node

In graphical depictions, we draw up-arrows or down-arrows to indicate spins. If there is an edge from node v_j to node v_i , with label $J_{ij} \in \{\pm 1\}$, we say that this edge is *consistent with the spin assignment* Σ provided that:

$$J_{ij}\sigma_i\sigma_j = 1$$

which is the same as saying that $J_{ij} = \sigma_i\sigma_j$, or that $\sigma_i = J_{ij}\sigma_j$. An equivalent formalism is that in which edges are labeled by “0” or “1,” instead of 1 and −1, respectively, and edge labels J_{ij} belong to the set $\{0,1\}$, in which case consistency is the property that $J_{ij} \oplus \sigma_i \oplus \sigma_j = 0$ (sum modulo two).

We will say that Σ is a *consistent spin assignment for the graph* G (or simply that G is consistent) if every edge of G is consistent with Σ . In other words, for any pair of vertices v_i and v_j , if there is a positive edge from node v_j to node v_i , then v_j and v_i must have the same spin, and if there is a negative edge connecting v_j to v_i , then v_j and v_i must have opposite spins. (If there is no edge from v_j to v_i , this requirement imposes no restriction on their spins.)

In order to decide whether a graph admits any consistent spin assignment, it is not necessary to actually test all the possible 2^n spin assignments. It is very easy to prove that there is a consistent assignment if and only if *every undirected loop in the graph G has a net positive sign*, that is to say, an even number, possibly zero, of negative arrows. Equivalently, any two (undirected) paths between two nodes must have the same net sign. By undirected loops or paths, we mean that one is allowed to transverse an edge either forward or backward. Graphs that satisfy this positive-loop property have been called *balanced* by Harary (1953). A proof that consistency and balancing are equivalent was given in Theorem 3 in Harary (1953); it is very simple and proceeds as follows. If a consistent assignment exists, then, for any undirected loop $v_{i_1}, \dots, v_{i_k} = v_{i_1}$ starting from and ending at the node v_{i_1} , inductively one has that:

$$\sigma_{i_1} = Q_{i_1, i_{k-1}} Q_{i_{k-1}, i_{k-2}} \dots Q_{i_2, i_1} \sigma_{i_1}$$

where $Q_{ij} = J_{ij}$ if we are transversing the edge from v_j to v_i , or $Q_{ij} = J_{ji}$ if we are transversing backward the edge from v_j to v_i . This implies (divide by σ_{i_1}) that the product of the edge signs is positive. Conversely, if any two paths between nodes have the same parity, and the graph is connected, pick node v_1 and label it “+” and then assign to every other node v_i the parity of a path connecting v_1 and v_i . (If the graph is not connected, do this construction on each component separately.)

The balancing property, in turn, can be checked with a fast dynamic programming-like algorithm. For connected graphs, there can be at most two consistent assignments, each of which is the reverse (flip every spin) of the other.

Monotone systems

A dynamical system is said to be *monotone* if there exists at least one consistent spin assignment for its associated graph G . Monotone systems (Smith 1995; Hirsch 1983, 1985) were introduced by Hirsch, and constitute a class of dynamical systems for which a rich theory exists. (To be precise, we have only defined the subclass of systems that are *monotone with respect to some orthant order*. The notion of monotonicity can be defined with respect to more general orders.)

Consistent response to perturbations

Monotonicity reflects the fact that a system responds consistently to perturbations on its components. Let us now discuss this property in informal terms. We view the nodes of the graph shown in Fig. 4a as corresponding to variables in the system, which quantify the concentrations of chemical species such as activated receptors, proteins, transcription factors, and so forth. Suppose that a perturbation, for example due to the external activation of a receptor represented by node 1, instantaneously increases the value of the concentration of this species. We represent this increase by an up-arrow inserted into that node, as in Fig. 4b. The effect on the other nodes is then completely predictable from the graph. The species associated to node 2 will decrease, because of the inhibiting character of the connection from 1 to 2, and the species associated to node 3 will increase (activating effect). Where monotonicity plays a role is in insuring that the concentration of the species corresponding to node 4 will also increase. It increases both because it is activated by 3, which has increased, and because it is inhibited by 2, so that less of 2 implies a smaller inhibition effect. Algebraically, the following expression involving partial derivatives:

$$\frac{\partial f_4}{\partial x_3} \frac{\partial f_3}{\partial x_1} + \frac{\partial f_4}{\partial x_2} \frac{\partial f_2}{\partial x_1}$$

(where f_i gives the rate of change of the i th species, in the differential equation model) is guaranteed to be positive, since it is a sum of positive terms: $(+)(+) + (-)(-)$. Intuitively, the expression measures the sensitivity of the

rate of change dx_4/dt of the concentration of 4 with respect to perturbations in 1, with the two terms giving the contributions for each of the two alternative paths from node 1 to node 4. *This unambiguous global effect holds true regardless of the actual values of parameters such as kinetic constants, and even the algebraic forms of reactions, and depends only on the signs of the entries of the Jacobian of f .* Observe that the arrows shown in Fig. 4b provide a consistent spin-assignment for the graph, so the system is monotone.

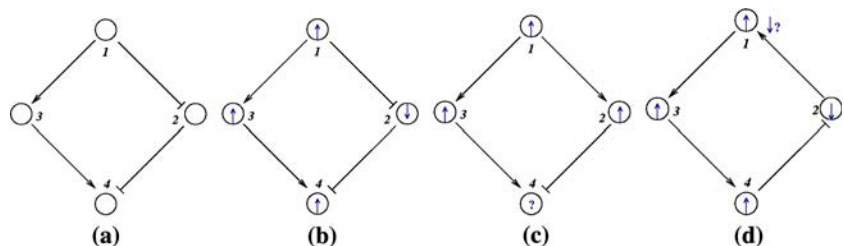
In contrast, consider next the graph in Fig. 4c, where the edge from 1 to 2 is now positive. There are two paths from node 1 to node 4, one of which (through 3) is positive and the other of which (through 2) is negative. Equivalently, the undirected loop 1,3,4,2,1 (“undirected” because the last two edges are transversed backward) has a net negative parity. Therefore, the loop test for consistency fails, so that there is no possible consistent spin-assignment for this graph, and therefore the corresponding dynamical system is not monotone. Reflecting this fact, the net effect of an increase in node 1 is ambiguous. *It is impossible to conclude from the graphical information alone whether node 4 will be repressed (because of the path through 2) or activated (because of the path through 3).* There is no way to resolve this ambiguity unless equations and precise parameter values are assigned to the arrows.

To take a concrete example, suppose that the equations for the system are as follows:

$$\frac{dx_1}{dt} = 0 \quad \frac{dx_2}{dt} = x_1 \quad \frac{dx_3}{dt} = x_1 \quad \frac{dx_4}{dt} = x_4(k_3x_3 - k_2x_2),$$

where the reaction constants k_2 and k_3 are two positive numbers. The initial conditions are taken to be $x_1(0) = x_4(0) = 1$, and $x_2(0) = x_3(0) = 0$, and we ask how the solution $x_4(t)$ will change when the initial value $x_1(0)$ is perturbed. With $x_1(0) = 1$, the solution is $x_4(t) = \exp \alpha t^2/2$, where $\alpha = k_3 - k_2$. On the other hand, if $x_1(0)$ is perturbed to a larger value, let us say $x_1(0) = 2$, then $x_4(t) = \exp \alpha t^2$. This new value of $x_4(t)$ is larger than the original unperturbed value $\exp \alpha t^2/2$ provided that $\alpha > 0$, but it is smaller than it if, instead, $\alpha < 0$. In other words, the sign of the sensitivity of x_4 to a perturbation on x_1 cannot be predicted from knowledge of the graph alone, but it depends on

Fig. 4 (a) and (b) graph and consistent assignment, (c) and (d) no possible consistent assignments



whether $k_2 < k_3$ or $k_2 > k_3$. Compare this with the monotone case, as in Fig. 4a. A concrete example is obtained if we modify the x_2 equation to $dx_2/dt = 1/(1+x_1)$. Now the solutions are $x_4(t) = \exp \beta_1 t^2$ and $x_4(t) = \exp \beta_2 t^2$, respectively, with $\beta_1 = k_3/2 - k_2/4$ and $\beta_2 = k_3 - k_2/6$, so we are guaranteed that x_4 is larger in the perturbed case, a conclusion that holds true no matter what are the numerical values of the (positive) constants k_i .

The uncertainty associated to a graph like the one in Fig. 4c might be undesirable in natural systems. Cells of the same type differ in concentrations of ATP, enzymes, and other chemicals, and this affects the values of model parameters, so two cells of the same type may well react differently to the same “stimulus” (increase in concentration of chemical 1). While such epigenetic diversity is sometimes desirable, it makes behavior less predictable and robust. From an evolutionary viewpoint, a “change in wiring” such as replacing the negative edge from 1 to 2 by a positive one (or, instead, perhaps introducing an additional inconsistent edge) could lead to unpredictable effects, and so the fitness of such a mutation may be harder to evaluate. In a monotone system, in contrast, a stimulus applied to a component is propagated in an unambiguous manner throughout the circuit, promoting a predictably consistent increase or consistent decrease in the concentrations of all other components.

Similarly, consistency also applies to feedback loops. For example, consider the graph shown in Fig. 4d. The negative feedback given by the inconsistent path 1,3,4,2,1 means that the instantaneous effect of an up-perturbation of node 1 feeds back into a negative effect on node 1, while a down-perturbation feeds back as a positive effect. In other words, the feedback loop acts against the perturbation.

Of course, negative feedback as well as inconsistent feedforward circuits are important components of biomolecular networks, playing a major role in homeostasis and in signal detection. The point being made here is that inconsistent networks may require a more delicate tuning in order to perform their functions.

In rigorous mathematical terms, this predictability property can be formulated as *Kamke’s Theorem*. Suppose that $\Sigma = \{\sigma_i, i = 1, \dots, n\}$ is a consistent spin assignment for the system graph G . Let $x(t)$ be any solution of

$dx/dt = f(x)$. We wish to study how the solution $z(t)$ arising from a perturbed initial condition $z(0) = x(0) + \Delta$ compares to the solution $x(t)$. Specifically, suppose that a positive perturbation is performed at time $t = 0$ on the i th coordinate, for some index $i \in \{1, \dots, n\}$: $z_i(0) > x_i(0)$ and $z_j(0) = x_j(0)$ for all $j \neq i$. For concreteness, let us assume that the perturbed node i has been labeled by $\sigma_i = +1$. Then, Kamke’s Theorem says the following: for each node that has the same parity (i.e., each index j such that $\sigma_j = +1$), and for every future time t , $z_j(t) \geq x_j(t)$. Similarly, for each node with opposite parity ($\sigma_j = -1$), and for every time t , $z_j(t) \leq x_j(t)$. (Moreover, one or more of these inequalities must be strict.) This is the precise sense in which an up-perturbation of the species represented by node v_i unambiguously propagates into up- or down-behavior of all the other species. See Smith (1995) for a proof, and see Angeli and Sontag (2003) for generalizations to systems with external input and output channels.

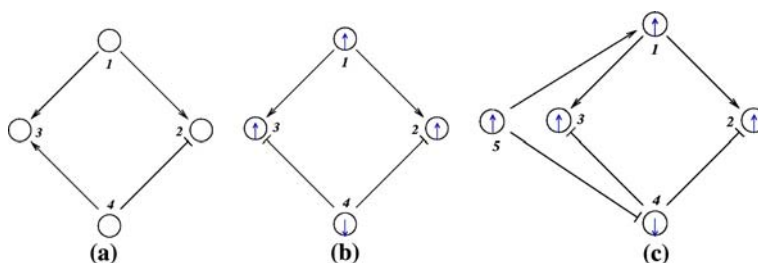
For difference equations (discrete time systems), once that self-loops have been included in the graph G and the definition of consistency, Kamke’s theorem also holds; in this case the proof is easy, by induction on time steps.

Consistent graphs can be embedded into larger consistent ones, but inconsistent ones cannot. For example, consider the graph shown in Fig. 5a. This graph admits no consistent spin assignment since the undirected loop 1,3,4,2,1 has a net negative parity. Thus, there cannot be any consistent graph that includes this graph as a subgraph. Compare this with the graph shown in Fig. 5b. Consistency of this graph may well represent consistency of a larger graph which involves a yet-undiscovered species, such as node 5 in Fig. 5c. Alternatively, and from an “incremental design” viewpoint, this graph being consistent makes it possible to consistently add node 5 in the future.

Removing the smallest number of edges so as to achieve consistency

Let us call the *consistency deficit (CD)* of a graph G the smallest possible number of edges that should be removed from G in order that there remains a consistent graph, and, correspondingly, a monotone system.

Fig. 5 (a) inconsistent, (b) consistent, (c) adding node to consistent network



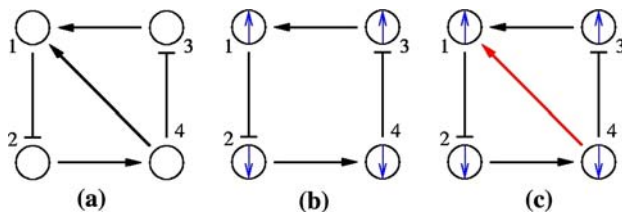


Fig. 6 (a) inconsistent graph, (b) consistent subgraph, (c) one inconsistent edge

As an example, take the graph shown in Fig. 6a. For this graph, it suffices to remove just one edge, the diagonal positive one, so the CD is 1. (In this example, the solution is unique, in that no other single other edge would suffice, but for other graphs there are typically several alternative ways to achieve consistency with a minimal number of deletions.)

After deleting the diagonal, a consistent spin assignment Σ is: $\sigma_1 = \sigma_3 = 1$ and $\sigma_2 = \sigma_4 = -1$, see Fig. 6b. (Another assignment is the one with all spins reversed: $\sigma_1 = \sigma_2 = -1$ and $\sigma_3 = \sigma_4 = 1$.) If we now bring back the deleted edge, we see that in the original graph only the one edge from node 1 to node 4 is inconsistent for the spin assignment Σ (Fig. 6c).

This example illustrates a general fact: minimizing the number of edges that must be removed so that there remains a consistent graph is equivalent to finding a spin assignment Σ for which the number of inconsistent edges (those for which $J_{ij}\sigma_i\sigma_j = -1$) is minimized.

Yet another rephrasing is as follows. For any spin assignment Σ , let A_1 be the subset of nodes labeled +1, and let A_{-1} be the subset of nodes labeled -1. The set of all nodes is partitioned into A_1 and A_{-1} . (In Fig. 6b, we have $A_1 = \{1, 3\}$ and $A_{-1} = \{2, 4\}$.) Conversely, any partition of the set of nodes into two subsets can be thought of as a spin assignment. With this interpretation, a consistent spin assignment is the same as a partition of the node set into two subsets A_1 and A_{-1} in such a manner that all edges between elements of A_1 are positive, all edges between elements of A_{-1} are positive, and all edges between a node in A_1 and a node in A_{-1} are negative, see Fig. 7. (A sociological interpretation of these partitions motivated the original paper (Harary 1953): vertices represent people, edges their likes and dislikes of each other, and consistency or balancing means that one may partition the people (nodes) into two cohesive groups that dislike each other.) More generally, computing the CD amounts to finding a partition so that $n_1 + n_{-1} + p$ is minimized, where n_1 is the number of negative edges between nodes in A_1 , n_{-1} is the number of negative edges between nodes in A_{-1} , and p is the number of positive edges between nodes in A_1 and A_{-1} .

A very special case is when the graph has all of its edges labeled negative, that is, $J_{ij} = -1$ for all i, j . Stated in the

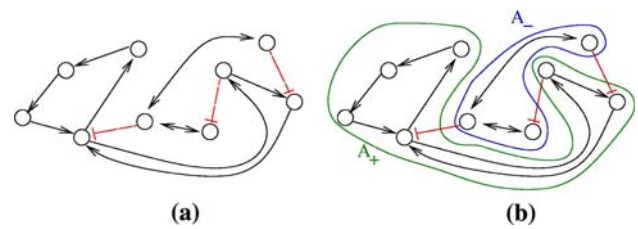


Fig. 7 (a) Consistent graph; (b) partition into A_1 and A_{-1}

language of partitions, the CD problem amounts to searching for a partition such that $n_1 + n_{-1}$ is minimized (as there are no positive edges, $p = 0$). Moreover, since there are no positive edges, $n_1 + n_{-1}$ is actually the total number of edges between any two nodes in A_1 or in A_{-1} . Thus, $N - (n_1 + n_{-1})$ is the number of remaining edges, that is, the number of edges between nodes in A_1 and A_{-1} . Therefore, minimizing $n_1 + n_{-1}$ is the same as maximizing $N - (n_1 + n_{-1})$. This is precisely the standard “MAX-CUT” problem in computer science.

As a matter of fact, not only is MAX-CUT a particular case, but, conversely, it is possible to reduce the CD problem to MAX-CUT by means of the following trick. For each edge labeled +1, say from v_i to v_j , delete the edge but insert a new node w_{ij} , and two negative edges, one from v_i to w_{ij} and one from w_{ij} to v_j :

$$v_i \rightarrow v_j \rightsquigarrow v_i \dashv w_{ij} \dashv v_j.$$

The enlarged graph has only negative edges, and it is easy to see that the minimal number of edges that have to be removed in order to achieve consistency is the same as the number of edges that would have had to be removed in the original graph. Unfortunately, the MAX-CUT problem is NP-hard. However, the paper (DasGupta et al. 2007) gave an approximation polynomial-time algorithm for the CD problem, guaranteed to solve the problem to within 87.9% of the optimum value, as an adaptation of the semi-definite programming relaxation approach to MAX-CUT based on Goemans and Williamson’s work (1995). (Is not enough to simply apply the MAX-CUT algorithm to the enlarged graph obtained by the above trick, because the approximation bound is degraded by the additional edges, so the construction takes some care.) The recent paper (Hüffner et al. 2007) substantially improved upon the approach in DasGupta et al. (2007), resulting in a very efficient algorithm.

Relation to Ising spin-glass models

Another interpretation of CD uses the language of statistical mechanics. An *Ising spin-glass model* is defined by a graph G together with an “interaction energy” J_{ij}

associated to each edge (in our conventions, J_{ij} is associated to the edge from v_j to v_i). In binary models, $J_{ij} \in \{1, -1\}$, as we have here. A spin-assignment Σ is also called a (magnetic) “spin configuration.” A “non-frustrated” spin-glass model is one for which there is a spin configuration for which every edge is consistent (Barahona 1982; De Simone et al. 1995; Istrail 2000). This is the same as a consistent assignment for the graph G in our terminology. Moreover, a spin configuration that maximizes the number of consistent edges is one for which the “free energy” (with no exterior magnetic field):

$$H(\Sigma) = - \sum_{ij} J_{ij} \sigma_i \sigma_j$$

is minimized. This is because, if Σ results in $C(\Sigma)$ consistent edges, then $H(\Sigma) = -C(\Sigma) + I(\Sigma) = T - 2C(\Sigma)$, where $I(\Sigma)$ is the number of non-consistent edges for the assignment Σ and $T = C + I$ is the total number of edges; thus, minimizing $H(\Sigma)$ is the same as maximizing $C(\Sigma)$. A minimizing Σ is called a “ground state.” (A special case is that in which $J_{ij} = -1$ for all edges, the “anti-ferromagnetic case.” This is the same as the MAX-CUT problem.)

Near-monotone systems may be “practically” monotone

Obviously, there is no reason for large biochemical networks to be consistent, and they are not. However, when the number of inconsistencies in a biological interaction graph is small, it may well be the case that the network is in fact consistent in a practical sense. For example, a gene regulatory network represents all *potential* effects among genes. These effects are often mediated by proteins which themselves need to be activated in order to perform their function, and this activation will, in turn, be contingent on the “environmental” context: extracellular ligands, additional genes being expressed which may depend on cell type or developmental stage, and so forth. Thus, depending on the context, different subgraphs of the original graph describe the system, and these graphs may be individually consistent even if the entire graph, the union of all these subgraphs, is not. As an illustration, take the system in Fig. 4c. Suppose that under environmental conditions A, the edge from 1 to 2 is not present, and under non-overlapping conditions B, the edge from 1 to 3 is not be present. Then, under either conditions, A or B, the graph is consistent, even though, formally speaking, the entire network is not consistent.

The closer to consistent, the more likely that this phenomenon may occur.

Some evidence suggesting near-monotonicity of natural networks

Since consistency in biological networks may be desirable, one might conjecture that natural biological networks tend to be consistent. As a way to test this hypothesis, the CD algorithm from DasGupta et al. (2007) was run on the yeast *Saccharomyces cerevisiae* gene regulatory network from Milo et al. (2002), downloaded from <http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/yeast-Data.mat> (Milo et al. (2002) used the YPD database (Costanzo et al. 2001). Nodes represent genes, and edges are directed from transcription factors, or protein complexes of transcription factors, into the genes regulated by them.) This network has 690 nodes and 1,082 edges, of which 221 are negative and 861 are positive (we labeled the one “neutral” edge as positive; the conclusions do not change substantially if we label it negative instead, or if we delete this one edge). The approximation algorithm from DasGupta et al. (2007) estimated the CD at 43, and the exact algorithm from Hüffner et al. (2007) later improved this estimate to a precise value $CD = 41$. In other words, *deleting a mere 4% of edges makes the network consistent*. Also remarkable is the following fact. The original graph has 11 components: a large one of size 664, one of size 5, three of size 3, and six of size 2. All of these components *remain connected* after edge deletion. The deleted edges are all from the largest component, and they are incident on a total of 65 nodes in this component.

To better appreciate if a small CD might happen by chance, the algorithm was also run on random graphs having 690 nodes and 1082 edges (chosen uniformly), of which 221 edges (chosen uniformly) are negative. It was found that, for such random graphs, about 12.6% (136.6 ± 5) of edges have to be removed in order to achieve consistency. (To analyze the scaling of this estimate, we generated random graphs with N nodes and $1.57N$ edges of which $0.32N$ are negative. We found that for $N > 10$, approximately $N/5$ nodes must be removed, thus confirming the result for $N = 690$.) Thus, the CD of the biological network is roughly 15 standard deviations away from the mean for random graphs. Both topology (i.e., the underlying graph) and actual signs of edges contribute to this near-consistency of the yeast network. To justify this assertion, the following numerical experiment was performed. We randomly changed the signs of 50 positive and 50 negative edges, thus obtaining a network that has the same number of positive and negative edges, and the same underlying graph, as the original yeast network, but with 100 edges, picked randomly, having different signs. Now, one needs 8.2% (88.3 ± 7.1) deletions, an amount in-between that obtained for the original yeast network and the one obtained for random graphs. Changing more signs, 100

positives and 100 negatives, leads to a less consistent network, with 115.4 ± 4.0 required deletions, or about 10.7% of the original edges, although still not as many as for a random network.

Decomposing systems into monotone components

Another motivation for the study of near-monotone systems is from decomposition-based methods for the analysis of systems that are interconnections of monotone subsystems. One may “pull out” inconsistent connections among monotone components, in such a manner that the original system can then be viewed as a “negative feedback” loop around an otherwise consistent system (Fig. 8). In this interpretation, the number of interconnections among monotone components corresponds to the number of variables being fed-back.

For example, let us take the graph shown in Fig. 6a. The procedure of dropping the diagonal edge and seeing it instead as an external feedback loop can be modeled as follows. The original differential equation $dx_1/dt = f_1(x_1, x_2, x_3, x_4)$ is replaced by the equation $dx_1/dt = f_1(x_1, x_2, x_3, u)$, where the symbol u , which represents an external input signal, is inserted instead of the state variable x_4 . The consistent system in Fig. 8 includes the remaining four edges, and the “negative” feedback (negative in the sense that it is inconsistent with the rest of the system) is the connection from x_4 , seen as an “output” variable, back into the input channel represented by u . The closed-loop system obtained by using this feedback is the original system, now viewed as a negative feedback around the consistent system in Fig. 6b.

Generally speaking, the decomposition techniques in Angeli and Sontag (2003, 2004a), Angeli et al. (2004a, b), Sontag (2004, 2005), Enciso et al. (2006), de Leenheer et al. (2005), Enciso and Sontag (2005b, 2006), De Leenheer and Malisoff (2006), Gedeon and Sontag (2007) are most useful if the feedback loop involves few variables. This is equivalent to asking that the graph G associated to the system be close to consistent, in the sense of the CD of G being small. This view of systems as monotone systems—which have strong stability properties, as discussed next, with negative-feedback regulatory loops around them is very appealing from a control engineering perspective as well.

Dynamical behavior of monotone systems

Continuous-time monotone systems have convergent behavior. For example, they cannot admit any possible stable oscillations (Hirsch and Smith 2005; Haderl and Glas 1983; Hirsch 1984). When there is only one

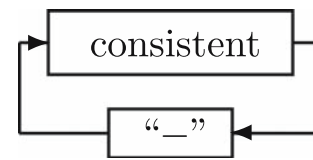


Fig. 8 Pulling-out inconsistent connections

steady-state, a theorem of Dancer (1998) shows—under mild assumptions regarding possible constraints on the values of the variables, which are often satisfied, and boundedness of solutions, which usually follows from conservation laws—that every solution converges to this unique steady-state (monostability). When, instead, there are multiple steady-states, the Hirsch Generic Convergence Theorem (Smith 1995; Hirsch and Smith 2005; Hirsch 1983, 1985) is the fundamental result. A *strongly* monotone system is one for which the an initial perturbation $z_i(0) > x_i(0)$ on the concentration of any species propagates as a strict up or down perturbation: $z_j(t) > x_j(t)$ for all $t > 0$ and all indices j for which $\sigma_j = \sigma_i$, and $z_j(t) < x_j(t)$ for all $t > 0$ and all j for which $\sigma_j = -\sigma_i$. Observe that this requirement is stronger (hence the terminology) than merely weak inequalities: $z_j(t) \geq x_j(t)$ or $z_j(t) \leq x_j(t)$, respectively as in Kamke’s Theorem. A sufficient condition for strong monotonicity is that the Jacobian matrices must be irreducible for all x , which basically amounts to asking that the graph G must be strongly connected and that every non-identically zero Jacobian entry be everywhere non-zero. Even though they may have arbitrarily large dimensionality, monotone systems behave in many ways like one-dimensional systems: Hirsch’s Theorem asserts that generic bounded solutions of strongly monotone differential equation systems must converge to the set of steady-states. (“Generic” means “every solution except for a measure-zero set of initial conditions.”) In particular, no “chaotic” or other “strange” dynamics can occur. For discrete-time strongly monotone systems, generically also stable oscillations are allowed besides convergence to equilibria, but no more complicated behavior.

The ordered behavior of monotone systems is robust with respect to spatial localization effects as well as signaling delays (such as those arising from transport, transcription, or translation). Moreover, their stability character does not change much if some inconsistent connections are inserted, but only provided that these added connections are weak (“small gain theorem”) or that they operate at a comparatively fast time scale (Wang and Sontag 2006a).

The intuition behind the convergence results is easy to explain in the very special case of just two interacting species, described by a two-dimensional system with variables $x(t)$ and $y(t)$:

$$\frac{dx}{dt} = f(x, y)$$

$$\frac{dy}{dt} = g(x, y).$$

A system like this is monotone if either (a) the species are mutually activating (or, as is said in mathematical biology, “cooperative”), (b) they are mutually inhibiting (“competitive”), or (c) either x does not affect y , y does not affect x , or neither affects the other. Let us discuss the mutually activating case (a). (Case (b) is similar, and case (c) is easy, since the systems are partially or totally decoupled.) We want to argue that there cannot be any periodic orbit. Suppose that there would be a periodic orbit in which the motion is counterclockwise, as shown in Fig. 9a. We then pick two points in this orbit with identical x coordinates, as indicated by (x, y) and (x, y') in Fig. 9a. These points correspond to the concentrations at two times t_0, t_1 , with $x(t_0) = x(t_1)$ and $y(t_0) < y(t_1)$. Since $y(t_1)$ is larger than $y(t_0)$, x is at the same concentration, and the species are mutually activating, it follows that the rate of change in the concentration x should be comparatively larger at time t_1 than at time t_0 , that is, $f(x, y') \geq f(x, y)$. However, this contradicts the fact that $x(t)$ is increasing at time t_0 ($f(x, y) \geq 0$) but is decreasing at time t_1 ($f(x, y') \leq 0$). The contradiction means that there cannot be any counterclockwise-oriented curve. To show that there cannot be any clockwise-oriented curve, one may proceed by an entirely analogous argument, using two points (x, y) and (x', y) as in Fig. 9b. Of course, the power of monotone systems theory arises in the analysis of systems of higher dimension, since two-dimensional systems are easy to study by elementary phase plane methods.

For general, non-monotone systems, on the other hand, no dynamical behavior, including chaos, can be mathematically ruled out. This is in spite of the fact that some features of non-monotone systems are commonly regarded as having a stabilizing effect. For example, negative feedback loops confer robustness with regard to certain types of structural as well as external perturbations (Doyle

et al. 1990; Sepulchre et al. 1997; Sontag 1999; Khalil 2002). However, and perhaps paradoxically, the behavior of non-monotone systems may also be very fragile: for instance, they can be destabilized by delays in negative feedback paths. Nonetheless, we conjecture that systems that are close to monotone must be better-behaved, generically, than those that are far from monotone. Preliminary evidence (unpublished) for this has been obtained from the analysis of random Boolean networks, at least for discrete analogs of the continuous system, but the work is not yet definitive.

Directed cycles

Intuition suggests that somewhat less than monotonicity should suffice for guaranteeing that no chaotic behavior may arise, or even that no stable limit cycles exist. Indeed, monotonicity amounts to requiring that no undirected negative-parity cycles be present in the graph, but a weaker condition, that no *directed* negative parity cycles exist, should be sufficient to insure these properties. For a strongly connected graph, the property that no directed negative cycles exist is equivalent to the property that no undirected negative cycles exist, because the same proof as given earlier, but applied to directed paths, insures that a consistent spin assignment exists (and hence there cannot be any undirected negative cycles). However, for non-strongly connected graphs, the properties are not the same. On the other hand, every graph can be decomposed as a cascade of graphs that are strongly connected. This means (aside from some technicalities having to do with Jacobian entries being not identically zero but vanishing on large sets) that systems having no directed negative cycles can be written as a cascade of strongly monotone systems. Therefore, it is natural to conjecture that such cascades have nice dynamical properties. Indeed, under appropriate technical conditions for the systems in the cascade, one may recursively prove convergence to equilibria in each component, appealing to the theory of asymptotically autonomous systems (Thieme 1992) and thus one may conclude global convergence of the entire system (Hirsch 1989; Smith 1991). For example, a cascade of the form $dx/dt = f(x)$, $dy/dt = g(x, y)$ where the x system is monotone and where the system $dy/dt = g(x_0, y)$ is monotone for each fixed x_0 , cannot have any attractive periodic orbits (except equilibria). This is because the projection of such an orbit on the first system must be a point x_0 , and hence the orbit must have the form $(x_0, y(t))$. Therefore, it is an attractive periodic orbit of $dy/dt = g(x_0, y)$, and by monotonicity of this latter system we conclude that $y(t) \equiv$ a constant as well. The argument generalizes to any cascade, by an inductive

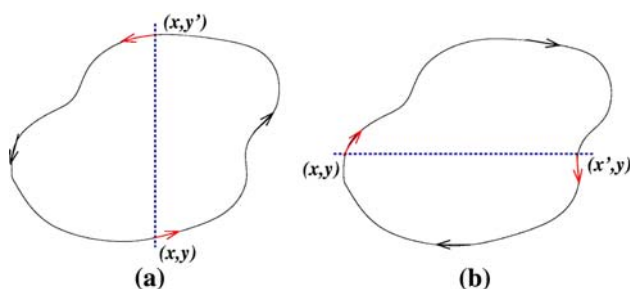


Fig. 9 Impossible (a) counterclockwise and (b) clockwise periodic orbits in planar cooperative system, each drawn in the (x, y) -plane

argument. Also, chaotic attractors cannot exist (D. Angeli et al. in preparation).

The condition of having no directed negative cycles is the weakest one that can be given strictly on the basis of the graph G , because for any graph G with a negative feedback loop there is a system with graph G which admits stable periodic orbits. (First find a limit cycle for the loop, and then use a small perturbation to define a system with nonzero entries as needed, which will still have a limit cycle.)

Positive feedback and stability

The strong global convergence properties of monotone systems mentioned above would seemingly contradict the fact that positive feedback, which tends to increase the direction of perturbations, is allowed in monotone systems, but negative feedback, which tends to stabilize systems, is not. One explanation for this apparent paradox is that the main theorems in monotone systems theory only guarantee that *bounded* solutions converge, but they do not make any assertions about unbounded solutions. For example, the system $dx/dt = -x + x^2$ has the property that every solution starting at an $x(0) > 1$ is unbounded, diverging to $+\infty$, a fact which does not contradict its monotonicity (every one-dimensional system is monotone). This is not as important a restriction as it may seem, because for biochemical systems it is often the case that all trajectories must remain bounded, due to conservation of mass and other constraints. A second explanation is that negative *self-loops* are not ruled out in monotone systems, and such loops, which represent degradation or decay diagonal terms, help insure stability.

Intuition on why negative self-loops do not affect monotonicity

In the definition of the graph associated to a continuous-time system, self-loops (diagonal terms in the Jacobian of the vector field f) were ignored. The theory (Kamke's condition) does not require self-loop information in order to guarantee monotonicity. Intuitively, the reason for this is that a larger initial value for a variable x_i implies a larger value for this variable, at least for short enough time periods, independently of the sign of the partial derivative df_i/dx_i (continuity of flow with respect to initial conditions). For example, consider a degradation equation $dp/dt = -p$, for the concentration $p(t)$ of a protein P. At any time t , we have that $p(t) = e^{-t}p(0)$, where $p(0)$ is the initial concentration. The concentration $p(t)$ is positively proportional to $p(0)$, even though the partial derivative $\partial(-p)/\partial p = -1$ is negative. Note that, in contrast, for a difference equation, a jump may occur: for instance the

iteration $p(t+1) = -p(t)$ has the property that the order of two elements is reversed at each time step. Thus, for difference equations, diagonal terms matter.

Multiple time scale analysis may make systems monotone

A system may fail to be monotone due to the effect of negative regulatory loops that operate at a faster time scale than monotone subsystems. In such a case, sometimes an approximate but monotone model may be obtained, by collapsing negative loops into self-loops. Mathematically: a non-monotone system might be a singular perturbation of a monotone system. A trivial linear example that illustrates this point is $dx/dt = -x - y$, $\varepsilon dy/dt = -y + x$, with $\varepsilon > 0$. This system is not monotone (with respect to any orthant cone). On the other hand, for $\varepsilon \ll 1$, the fast variable y tracks x , so the slow dynamics is well-approximated by $dx/dt = -2x$ (monotone, since every scalar system is). More generally, one may consider $dx/dt = f(x, y)$, $\varepsilon dy/dt = g(x, y)$ such that the fast system $dy/dt = g(x, y)$ has a unique globally asymptotically stable steady-state $y = h(x)$ for each x (and possibly a mild input to state stability requirement, as with the special case $\varepsilon dy/dt = -y + h(x)$), and the slow system $dx/dt = f(x, h(x))$ is (strongly) monotone. Then one may expect that the original system inherits global convergence properties, at least for all $\varepsilon > 0$ small enough. The paper (Wang and Sontag 2006b) employs tools from geometric invariant manifold theory (Fenichel 1979; Jones 1994), taking advantage of the existence of a manifold M_ε invariant for the dynamics, which attracts all near-enough solutions, and with an asymptotic phase property. The system restricted to the invariant manifold M_ε is a regular perturbation of the fast ($\varepsilon = 0$) system, and hence inherits strong monotonicity properties. So, solutions in the manifold will be generally well-behaved, and asymptotic phase implies that solutions track solutions in M_ε , and hence also converge to equilibria if solutions on M_ε do. However, the technical details are delicate, because strong monotonicity only guarantees generic convergence, and one must show that the generic tracking solutions start from the “good” set of initial conditions, for generic solutions of the large system.

Discrete-time systems

As discussed, for autonomous differential equations monotonicity implies that stable periodic behaviors will not be observed, and moreover, under certain technical assumptions, all trajectories must converge to steady-states. This is not exactly true for difference equation models, but a variant does hold: for discrete-time monotone

systems, trajectories must converge to either steady-states or *periodic orbits*. In general, even the simplest difference equations may exhibit arbitrarily complicated (chaotic) behavior, as shown by the logistic iteration in one dimension $x(t+1) = kx(t)(1-x(t))$ for appropriate values of the parameter k (Devaney 1989). However, for monotone difference equations, a close analog of Hirsch's Generic Convergence Theorem is known. Specifically, suppose that the equations are point-dissipative, meaning that all solutions converge to a bounded set (Hale 1988), and that the system is strongly monotone, in the sense that the Jacobian matrix $(\partial f_i / \partial x_j)$ is irreducible at all states. Then, a result of Tereščák and coworkers (Poláčik and Tereščák 1992; Poláčik and Tereščák 1993; Hess and Poláčik 1993; Tereščák 1996) shows that there is a positive integer m such that generic solutions (in an appropriate sense of genericity) converge to periodic orbits with period at most m . Results also exist under less than strong monotonicity, just as in the continuous case, for example when steady-states are unique (Dancer 1998).

Difference equations allow one to study wider classes of systems. As a simple example, consider the nondimensionalized harmonic oscillator (idealized mass-spring system with no damping), which has equations

$$\begin{aligned}\frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -x.\end{aligned}$$

(For this example, we allow variables to be negative; these variables might indicate deviations of concentrations from some reference value.) This system is not monotone, since $v_1 \rightarrow v_2$ is negative and $v_1 \rightarrow v_2$ is positive, so that its graph has a negative loop. On the other hand, suppose that one looks at this system every Δt seconds, where $\Delta t = \pi$. The discrete-time system that results (using a superscript $+$ to indicate time-stepping) is now:

$$\begin{aligned}x^+ &= -y \\ y^+ &= -x\end{aligned}$$

(this is obtained by solving the differential equation on an interval of length π). This system is monotone (both $v_1 \rightarrow v_2$ and $v_1 \rightarrow v_2$ are negative). Every trajectory of this discrete system is, in fact, of period two: $(x_0, y_0) \rightarrow (-y_0, -x_0) \rightarrow (x_0, y_0) \rightarrow \dots$. This periodic property for the difference equation corresponds to the period- 2π behavior of the original differential equation.

Oscillatory behaviors

Stable periodic behaviors are ruled-out in autonomous monotone continuous-time systems. However, stable

periodic orbits may arise through various external mechanisms. Three examples are (1) inhibitory negative feedback from some species into others in a monotone monostable system, (2) the generation of relaxation oscillations from a hysteresis parametric behavior by negative feedback on parameters by species in a monotone system, and (3) entrainment of external periodic signals. These general mechanisms are classical and well-understood for simple, one or two-dimensional, dynamics, and they may be generalized to the case where the underlying system is higher-dimensional but monotone.

Embeddings in monotone systems

As observed by Gouzé (1988), Gouze and Haderler (1994), any n -dimensional system can be viewed as a subsystem of a $2n$ -dimensional monotone system. The mathematical trick is to first duplicate every variable (species), introducing a “dual” species, and then to replace every inconsistent edge by an edge connecting the source species and the “dual” of its target (and vice-versa). The construction is illustrated in Fig. 10. At first, this embedding result may seem paradoxical, since all monotone (or strongly monotone) systems have especially nice dynamical behaviors, such as not having any attractive periodic orbits or chaotic attractors, and of course non-monotone systems may admit such behaviors. However, there is no contradiction. A non-monotone subsystem of a monotone system may well have, say, a chaotic attractor or a stable periodic orbit: it is just that this attractor or orbit will be unstable when seen as a subset of the extended ($2n$ -dimensional) state space. Not only there is no contradiction, but a classical construction of Smale (1976) shows that indeed any possible dynamics can be embedded in a larger monotone system. More generally, the Hirsch Generic Convergence Theorem guarantees convergence to equilibria from *almost every* initial condition; applied to the above construction, in general the exceptional set of initial conditions would include the “thin” set corresponding to the embedded subsystem. Yet, one may ask what happens for example if the larger $2n$ -system has a unique equilibrium. In that case, it is known (Dancer 1998) that every trajectory converges (not

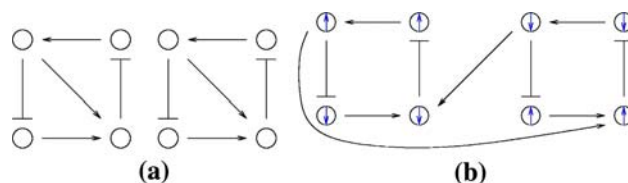


Fig. 10 (a) Duplicated inconsistent graph, (b) replacing arrows and consistent assignment

merely generic ones), so, in particular, the embedded subsystem must also be “well-behaved.” Thus, systems that may be embedded by the above trick into monotone systems with unique equilibria will have global convergence to equilibria. This property amounts to the “small gain theorem” shown in Angeli and Sontag (2003), see Enciso et al. (2006) for a discussion and further results using this embedding idea.

Discrete systems

We remark that one may also study difference equations for which the state components are only allowed to take values out of a finite set. For example, in Boolean models of biological networks, each variable $x_i(t)$ can only attain two values (0/1 or “on/off”). These values represent whether the i th gene is being expressed, or the concentration of the i th protein is above certain threshold, at time t . When detailed information on kinetic rates of protein–DNA or protein–protein interactions is lacking, and especially if regulatory relationships are strongly sigmoidal, such models are useful in theoretical analysis, because they serve to focus attention on the basic dynamical characteristics while ignoring specifics of reaction mechanisms (Kauffman 1969a, b; Kauffman and Glass 1973; Albert and Othmer 2003; Chaves et al. 2005).

For difference equations over finite sets, such as Boolean systems, it is quite clear that all trajectories must either settle into equilibria or to periodic orbits, whether the system is monotone or not. However, cycles in discrete systems may be arbitrarily long and these might be seen as “chaotic” motions. Monotone systems, while also settling into steady-states or periodic orbits, have generally shorter cycles. This is because periodic orbits must be anti-chains, i.e. no two different states can be compared; see Smith (1995) and Gilbert (1954). For example, consider a discrete-time system in which species concentrations are quantized to the k values $\{0, \dots, k-1\}$; we interpret monotonicity with respect to the partial order: $(a_1, \dots, a_n) \leq (b_1, \dots, b_n)$ if every coordinate $a_i \leq b_i$. For non-monotone systems, orbits can have as many as k^n states. On the other hand, monotone systems cannot have orbits of size more than the width (size of largest anti-chains) of $P = \{0, \dots, k-1\}^n$, which can be interpreted as the set of multisubsets of an n -element set, or equivalently as the set of divisors of a number of the form $(p_1 p_2 \dots p_n)^{k-1}$ where the p_i 's are distinct primes. The width of P is the number of possible vectors (i_1, \dots, i_n) such that $\sum i_j = \lfloor kn/2 \rfloor$ and each $i_j \in \{0, \dots, k-1\}$. This is a generalization of Sperner's Theorem; see Anderson (2002). For example, for $n = 2$, periodic orbits in a monotone system evolving on $\{0, \dots, k-1\}^2$ cannot have length larger than k , while non-monotone systems on

$\{0, \dots, k-1\}^2$ can have a periodic orbit of period k^2 . As another example, arbitrary Boolean systems (i.e., the state space is $\{0, 1\}^n$) can have orbits of period up to 2^n , but monotone systems cannot have orbits of size larger than $\binom{n}{\lfloor n/2 \rfloor} \approx 2^n \sqrt{2/(n\pi)}$. These are all classical facts in Boolean circuit design (Gilbert 1954). It is worth pointing out that any anti-chain P_0 can be seen as a periodic orbit of a monotone system. This is proved as follows: we enumerate the elements of P_0 as x_1, \dots, x_ℓ , and define $f(x_i) = x_{i-1}$ for all i modulo ℓ . Then, f can be extended to all elements of the state space by defining $f(x) = (0, \dots, 0)$ for every x which has the property that $x < x_i$ for some $x_i \in P_0$ and $f(x) = (k-1, \dots, k-1)$ for every x which is not $\leq x_i$ for any $x_i \in P_0$. It is easy to see that this is a monotone map (Gilbert 1954; Aracena et al. 2004).

While on the subject of discrete and in particular Boolean systems, we mention a puzzling fact: any Boolean function may be implemented by using just two inverters, with all other gates being monotone. In other words, a circuit computing any Boolean rule whatsoever may be built so that its “consistency deficit” is just two. This is a well-known fact in circuit design (Gilbert 1954; Minsky 1967). Here is one solution, from Clive (2006). One first shows how to implement the Boolean function that takes as inputs three bits A, B, C and outputs the vector of three complements ($\text{not}A, \text{not}B, \text{not}C$), by using this sequence of operations:

$$\begin{aligned} 2or3ones &= (A \wedge B) \vee (A \wedge C) \vee (B \wedge C) \\ 0or1ones &= \text{not}(2or3ones) \\ 1one &= 0or1ones \wedge (A \vee B \vee C) \\ 1or3ones &= 1one \vee (A \wedge B \wedge C) \\ 0or2ones &= \text{not}(1or3ones) \\ 0ones &= 0or2ones \wedge 0or1ones \\ 2ones &= 0or2ones \wedge 2or3ones \\ \text{not}A &= 0ones \vee (1one \wedge (B \vee C)) \vee (2ones \wedge (B \wedge C)) \\ \text{not}B &= 0ones \vee (1one \wedge (A \vee C)) \vee (2ones \wedge (A \wedge C)) \\ \text{not}C &= 0ones \vee (1one \wedge (A \vee B)) \vee (2ones \wedge (A \wedge B)) \end{aligned}$$

(the node labeled “2or3ones” computes the Boolean function “the input has exactly 2 or 3 ones” and so forth). Note that only two inverters have been used. If we now want to invert four bits A, B, C, D , we build the above circuit, but we implement the inversion of the three bits ($2or3ones, 1or3ones, D$) by a subcircuit with only two inverters. With a similar recursive construction, one may invert an arbitrary number of bits, using just two inverters.

I/O monotone systems

We next describe recent work on *monotone input/output systems* (“MIOS” from now on). Monotone i/o systems

originated in the analysis of mitogen-activated protein kinase cascades and other cell signaling networks, but later proved useful in the study of a broad variety of other biological models. Their surprising breadth of applicability notwithstanding, of course MIOS constitute a restricted class of models, especially when seen in the context of large biochemical networks. Indeed, the original motivation for introducing MIOS, in the 2003 paper (Angeli and Sontag 2003), was to study an existing *non-monotone* model of negative feedback in MAPK cascades. The key breakthrough was the realization that this example, and, as it turned out, many others, can be profitably studied by *decompositions into MIOS*. In other words, a non-monotone system is viewed as an interconnection of monotone subsystems. Based on the architecture of the interconnections between the subsystems (“network structure”), one deduces properties of the original, non-monotone, system. (Later work, starting with Angeli and Sontag (2004a), showed that even monotone systems can be usefully studied through this decomposition-based approach.)

We review the basic notion from Angeli and Sontag (2003). (For concreteness, we make definitions for systems of ordinary differential equations, but similar definitions can be given for abstract dynamical systems, including in particular reaction–diffusion partial differential equations and delay-differential systems, see e.g. Enciso and Sontag 2006) The basic setup is that of an input/output system in the sense of mathematical systems and control theory (Sontag 1998), that is, sets of equations

$$\frac{dx}{dt} = f(x, u), \quad y = h(x), \quad (1)$$

in which states $x(t)$ evolve on some subset $X \subseteq \mathbb{R}^n$, and input and output values $u(t)$ and $y(t)$ belong to subsets $U \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^p$, respectively. The coordinates x_1, \dots, x_n of states typically represent concentrations of chemical species, such as proteins, mRNA, or metabolites. The input variables, which can be seen as controls, forcing functions, or external signals, act as stimuli. Output variables can be thought of as describing responses, such as movement, or as measurements provided by biological reporter devices like GFP that allow a partial read-out of the system state vector (x_1, \dots, x_n) . The maps $f: X \times U \rightarrow \mathbb{R}^n$ and $h: X \rightarrow Y$ are taken to be continuously differentiable. (Much less can be assumed for many results, so long as local existence and uniqueness of solutions is guaranteed.) An *input* is a signal $u: [0, \infty) \rightarrow U$ which is locally essentially compact (meaning that images of restrictions to finite intervals are compact), and we write $\varphi(t, x_0, u)$ for the solution of the initial value problem $dx/dt(t) = f(x(t), u(t))$ with $x(0) = x_0$, or just $x(t)$ if x_0 and u are clear from the context, and $y(t) = h(x(t))$. See Sontag

(1998) for more on i/o systems. For simplicity of exposition, we make the blanket assumption that solutions do not blow-up on finite time, so $x(t)$ (and $y(t)$) are defined for all $t \geq 0$. (In biological problems, almost always conservation laws and/or boundedness of vector fields insure this property. In any event, extensions to local semiflows are possible as well.)

Given three partial orders on X, U, Y (we use the same symbol \preceq for all three orders), a monotone I/O system (MIOS), with respect to these partial orders, is a system (1000) such that h is a monotone map (it preserves order) and: for all initial states x_1, x_2 for all inputs u_1, u_2 , the following property holds: if $x_1 \preceq x_2$ and $u_1 \preceq u_2$ (meaning that $u_1(t) \preceq u_2(t)$ for all $t \geq 0$), then $\varphi(t, x_1, u) \preceq \varphi(t, x_2, u_2)$ for all $t > 0$. Here we consider partial orders induced by closed proper cones $K \subseteq \mathbb{R}^\ell$, in the sense that $x \preceq y$ iff $y - x \in K$. The cones K are assumed to have a nonempty interior and are pointed, i.e. $K \cap -K = \{0\}$. A *strongly* monotone system is one which satisfies the following stronger property: if $x_1 \preceq x_2$, $x_1 \neq x_2$, and $u_1 \preceq u_2$, then the strict inequality $\varphi(t, x_1, u) \prec \varphi(t, x_2, u_2)$ holds for all $t > 0$, where $x \prec y$ means that $y - x$ is in the interior of the cone K .

The most interesting particular case is that in which K is an *orthant* cone in \mathbb{R}^n , i.e. a set S_ε of the form $\{x \in \mathbb{R}^n \mid \varepsilon_i x_i \geq 0\}$, where $\varepsilon_i = \pm 1$ for each i .

When there are no inputs nor outputs, the definition of monotone systems reduces to the classical one of monotone dynamical systems studied by Hirsch, Smith, and Others (1995). This is what we discussed earlier, for the case of orthant cones. When there are no inputs, strongly monotone classical systems have especially nice dynamics. Not only is chaotic or other irregular behavior ruled out, but, in fact, almost all bounded trajectories converge to the set of steady states (Hirsch’s generic convergence theorem (see Hirsch (1983, 1985)).

A useful test for monotonicity with respect to orthant cones, which generalizes Kamke’s condition to the i/o case, is as follows. Let us assume that all the partial derivatives $\frac{\partial f_i}{\partial x_j}(x, u)$ for $i \neq j$, $\frac{\partial f_i}{\partial u_j}(x, u)$ for all i, j , and $\frac{\partial h_i}{\partial x_j}(x)$ for all i, j (subscripts indicate components) do not change sign, i.e., they are either always ≥ 0 or always ≤ 0 . We also assume that X is convex (much less is needed.) We then associate a directed graph G to the given MIOS, with $n + m + p$ nodes, and edges labeled “+” or “−” (or ± 1), whose labels are determined by the signs of the appropriate partial derivatives (ignoring diagonal elements of $\partial f / \partial x$). One may define in an obvious manner undirected loops in G , and the *parity* of a loop is defined by multiplication of signs along the loop. (See e.g. Angeli and Sontag 2004a, b for more details.) Then, it is easy to show that a system is monotone with respect to *some* orthant cones in X, U, Y if and only if there are no negative loops in G . A sufficient

condition for strong monotonicity is that, in addition to monotonicity, the partial Jacobians of f with respect to x should be everywhere irreducible. (“Almost-everywhere” often suffices; see Smith (1995), Hirsch and Smith 2005). See these references also for extensions to non-orthant cones in the case of no inputs and outputs, based on work of Schneider and Vidyasagar, Volkmann, and others (Schneider and Vidyasagar 1970; Volkmann 1972; Walcher 2001; Walter 1970).

In inhibitory feedback, a chemical species x_j typically affects the rate of formation of another species x_i through a term like $h(x_j) = V/(K + x_j)$. The decreasing function $h(x_j)$ can be seen as the output of an *anti-monotone* system, i.e. a system which satisfies the conditions for monotonicity, except that the output map *reverses* order: $x_1 \preceq x_2 \Rightarrow h(x_2) \succeq h(x_1)$.

An interconnection of monotone subsystems, that is to say, an entire system made up of monotone components, may or may not be monotone: “positive feedback” (in a sense that can be made precise) preserves monotonicity, while “negative feedback” destroys it. Thus, oscillators such as circadian rhythm generators require negative feedback loops in order for periodic orbits to arise, and hence are not themselves monotone systems, although they can be decomposed into monotone subsystems (cf. Angeli and Sontag 2004c). A rich theory is beginning to arise, characterizing the behavior of non-monotone interconnections. For example, Angeli and Sontag (2003) shows how to preserve convergence to equilibria; see also the follow-up papers (Enciso et al. 2006; Angeli et al. 2004b; de Leenheer et al. 2005; Enciso and Sontag 2006; Gedeon and Sontag 2007). Even for monotone interconnections, the decomposition approach is very useful, as it permits locating and characterizing the stability of steady-states based upon input/output behaviors of components, as described in Angeli and Sontag (2004a); see also the follow-up papers Angeli et al. (2004a), Enciso and Sontag (2005b), De Leenheer and Malisoff (2006).

Moreover, a key point brought up in Angeli and Sontag (2003, 2004a), Sontag (2004, 2005) is that new techniques for monotone systems in many situations allow one to characterize the behavior of an entire system, based upon the “qualitative” knowledge represented by general network topology and the inhibitory or activating character of interconnections, combined with only a relatively *small amount of quantitative* data. The latter data may consist of steady-state responses of components (dose-response curves and so forth), and there is no need to know the precise form of dynamics or parameters such as kinetic constants in order to obtain global stability conclusions and study global bifurcation behavior. We now discuss these issues, first for positive and then for negative feedback loops.

Positive feedback and possible multistability

We first discuss how multistability in cell signaling networks may arise from positive feedback loops. The general framework is that in which two input/output systems, each of which is monostable in isolation, can combine to produce a multi-stable closed-loop behavior when interconnected in closed-loop. Schematically, we consider two systems, one of which processes an input signal u and produces an output y , and a second one which processes the signal y to produce u .

The interconnection of these two systems is defined by feeding the output of each of the systems as an input to the other, Fig. 11. Steady-states of the closed-loop system correspond to those constant signals u and y that are obtained by intersecting the step-input steady-state responses (“characteristics” or “nonlinear DC gains,” defined below) of the individual systems. Such positive feedback systems may easily be multi-stable, even if the constituent pieces are monostable (Cinquin and Demongeot 2002; Thomas 1981; Snoussi 1998; Tyson et al. 2003). We next formally define characteristics, for any given system $dx/dt = f(x, u)$ with output $y = h(x)$.

Step-input steady-state responses (characteristics) of open-loop systems

For each constant input $u(t) \equiv u_0, t \geq 0$, we study the open-loop dynamical system $dx/dt = f(x, u_0)$ obtained by feeding this input. We will assume that all solutions approach steady-states, and we denote with $K(u_0)$ the set of possible steady-states, that is, the solutions x of the algebraic equation $f(x, u_0) = 0$. To each state x in this set $K(u_0)$, there is a corresponding output or measured quantity $h(x_0)$; we denote by $k(u_0)$ the set of all output values that arise in this manner. The graph of the set-valued mapping $u_0 \mapsto (u_0)$ is a subset of the cross product space $\mathbb{R}^m \times \mathbb{R}^p$, which may be thought of as a curve when $m = p = 1$, and which describes the possible steady-state output values for any given constant input. Although not strictly required, for simplicity we will assume from now on that these mappings are single-valued, not set-valued; in other words, that the open-loop system $dx/dt = f(x, u_0)$ is monostable, for any given constant level u_0 of the input. More precisely, we assume that a (single-valued) characteristic exists for the

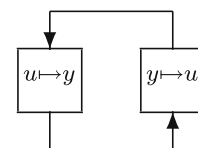


Fig. 11 Feedback interconnection of two systems

system: for each u_0 there is a unique steady-state for the dynamical system $dx/dt = f(x, u_0)$, denoted by $K(u_0)$, which is a globally asymptotically stable (“GAS”) steady-state. The (output) *characteristic* $k : U \rightarrow Y$ is then defined as the composition $h \circ K$. All solutions of $dx/dt = f(x, u_0)$ converge to $K(u_0)$, and the output $y(t)$ converges to $k(u_0)$, cf. Fig. 12. Another name for k is the *step-input steady-state response* or (nonlinear) *DC gain* of the system. In biological problems, a constant input may represent, for example, the concentration of a certain extracellular ligand in a signaling system, or the level of expression of a constitutively expressed gene.

Characteristics (dose–response curves, activity plots, steady-state expression of a gene in response to an external ligand, etc.) are frequently available from experimental data, especially in molecular biology and pharmacology, for instance in the modeling of receptor–ligand interactions (Chaves et al. 2004).

The results to be described are also valid under weaker definitions of characteristics, such as not requiring GAS properties, or allowing set-valued characteristics (Angeli and Sontag 2003; Angeli et al. 2004b; de Leenheer et al. 2005; De Leenheer and Malisoff 2006; Enciso and Sontag 2005a, 2006; G.A. Enciso and E.D. Sontag, in preparation).

It is worth pointing out that, if a system is monotone, then the stability property in the definition of characteristic is often automatically satisfied, provided that uniqueness of steady-states holds. More precisely, if one knows that (a) trajectories are bounded, and (b) the state space X has the property that least upper bounds and greatest lower bounds exist for any two elements of X (for example, if the state space is a “cube” with respect to the order cone K), then just knowing that $K(u_0)$ has only one point is enough to conclude that $K(u_0)$ is in fact a GAS state for $dx/dt = f(x, u_0)$ (Dancer 1998; Jiang 1994).

Hyperbolic and sigmoidal characteristics

Before reviewing theorems about feedback interconnections of MIOS systems, we discuss a very simple example which does not require any theory. Often, models of systems representing signaling and other molecular biology networks have a hyperbolic or a sigmoidal steady-state response.

To illustrate the first of these types of responses, we consider a protein P whose time-varying concentration $p(t)$ is subject to a Michaelis–Menten rate of production (initially linearly proportional to substrate concentration, but saturating at a maximal rate when a certain substrate U is in abundance), balanced by a linear rate of degradation/dilution. A simple model is as follows:

$$\frac{dp}{dt} = \frac{V_{\max}u}{(k_m + u)} - kp,$$

where $u = u(t)$ represents the concentration of the substrate U that is used in P’s formation. We view P itself as the output, that is $y(t) = p(t)$. The steady-state, when the input $u(t) \equiv u_0$ is constant, can be solved for by setting $dp/dt = 0$, from which we obtain:

$$p_0 = k(u_0) = \frac{(V_{\max}/k)u_0}{k_m + u_0}.$$

This is a *hyperbolic* response, Fig. 13a. The response is *graded* (“light-dimmer”): it is proportional to the parameter u_0 over a large range of values, until it saturates at a maximal level.

The second type, sigmoidal responses, arise from high-order phenomena, typically involving cooperativity. Suppose that $r > 1$ molecules of the substrate U are needed in order to produce a molecule of P. One usually models this situation by using a Hill rather than a Michaelis–Menten production rate:

$$\frac{dp}{dt} = \frac{V_{\max}u^r}{k_m^r + u^r} - kp$$

where $r > 1$ is a “Hill coefficient” or cooperativity index. (For $r = 1$, we have a Michaelis–Menten rate.) A *sigmoidal* (“doorbell”) steady-state response

$$p_0 = k(u_0) = \frac{(V_{\max}/k)u_0^r}{k_m^r + u_0^r}$$

results, see Fig. 13b. There is an inflection point of the graph at the value $u_0 = k_m$, and the plot becomes more steep and closer to a step function (with a switch at k_m) for larger r . Roughly speaking, values of the input $u_0 < k_m$ will

Fig. 12 Characteristics: (a) constant input, (b) convergence of internal states, (c) convergence of output to value $k(u)$

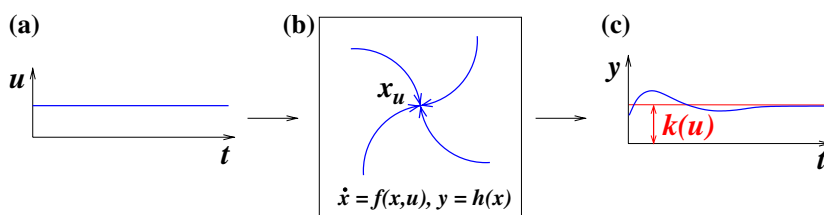
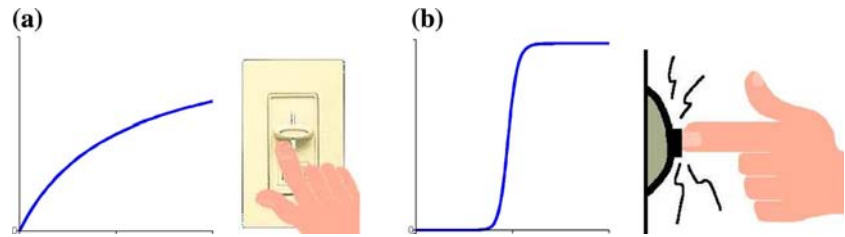


Fig. 13 (a) Hyperbolic response (b) sigmoidal response



not result in an appreciable activity of P ($p_0 \approx 0$ in steady-state) and values $u_0 > k_m$ result in a maximal value ($p_0 \approx V_{max}/k$ in steady-state).

It is believed that sigmoidal responses in signaling pathways are used in those situations in which binary decisions must be taken, such as when a cell must “decide” whether a gene should be transcribed or not, depending on the value of an extracellular signal (Novic and Weiner 1957; Ptashne 1992; Thomas and Kaufman 2001; Sha et al. 2003; Pomerening et al. 2003; Ferrell and Xiong 2001; Lisman 1985; Laurent and Kellershohn 1999; Gardner et al. 2000; Ferrell and Machleder 1998; Bagowski and Ferrell Jr. 2001; Bhalla et al. 2002; Cross et al. 2002; Becskei et al. 2001; Bagowski et al. 2003). Sigmoidal responses with large $r > 1$ (“ultrasensitive responses”) can be obtained by cascading simple enzymatic reactions provided that each reaction in the cascade has a Hill coefficient $r > 1$ (Ferrell Jr. 1996). (Basically, this statement amounts to the chain rule for derivatives.)

Creating bistability from sigmoidal responses

The simplest way to create bistability from a sigmoidal response is through positive feedback. We illustrate this procedure using the example just discussed. Schematically, we start with the “open loop” system that produces the protein P, with its concentration $y(t) = p(t)$ considered as an output and the concentration $u(t)$ of U seen as an input. We then “close the loop” by introducing a second system, one that simply produces U from P in such a manner that the concentration of U is proportional to that of P, as in Fig. 11. We ignore, for the purpose of this expository example, the details of the mechanism that implements the autocatalytic process in which U is produced from P. The mechanism might involve several intermediate proteins as well as time delays. For simplicity, we assume that there results an instantaneous change in the concentration of U proportional to the concentration of P. (One of the tools to be discussed, the theory of monotone input/output systems, provides conditions that explain when this simplification is justified.)

Mathematically, we simply replace the term u in the equation for dp/dt by λp , where the constant λ may be

thought of as a feedback gain. Absorbing the factor λ into V_{max} and k_m , we have the following equation:

$$\frac{dp}{dt} = (V_{max}p^r)/(k_m^r + p^r) - kp.$$

We plot in Fig. 14 both the formation rate $(V_{max}p^r)/(k_m^r + p^r)$ together with the degradation/dilution rate kp , in cases where $r = 1$ (left) or $r > 1$ (right). We assume, in the sigmoidal case, that the slope of the degradation curve is so that three intersections result, as shown in the plots. (For different k 's, the line will have different slopes, and anywhere from one to three intersections are possible.)

The behavior of solutions is clear from these graphs. In the case of hyperbolic responses, corresponding to $r = 1$, we see that for small p the formation rate is larger than the degradation rate, but for large p the opposite holds. Therefore, the concentration $p(t)$ converges to a unique intermediate value. A monostable closed-loop system results. In the sigmoidal case $r > 1$ (assuming three intersections), on the other hand, for p small the degradation rate is larger than the formation rate, so that $p(t)$ converges to a low value; on the other hand, for large p the formation rate is larger than degradation, and thus $p(t)$ converges to a high value instead. Thus, two stable states are created, one low and one high, by this interaction of formation and degradation. (There is also an intermediate, unstable state.) The reasoning followed above is totally elementary, but it serves to provide an intuition for the monotone approach (Angeli and Sontag 2004a), which may be seen as a far-reaching generalization of this reasoning. (Previous, more restricted, generalizations, were obtained in Rapp (1975), Hastings et al. (1977), Tyson and Othmer (1978), Smith (1987, 1995), Allwright (1977), Othmer (1976), Thron (1991), Mallet-Paret and Smith (1990) and Gedeon (1998)).

Positive feedback and multistability in monotone I/O systems

The elementary and intuitive proof of bistability for the simple production/degradation system with sigmoidal characteristics just discussed can be generalized to a

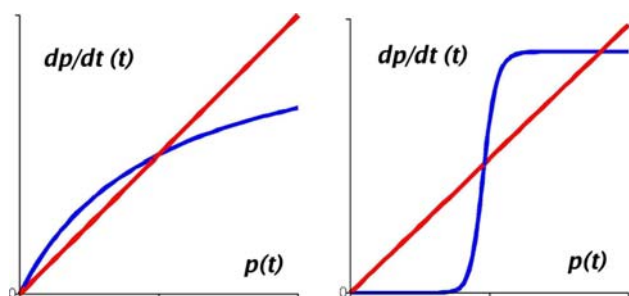


Fig. 14 Intersections of hyperbolic and sigmoidal response with degradation

feedback interconnections of individually monostable systems, applying even if the $u \mapsto$ and $y \mapsto$ systems in Fig. 11 are far more complicated than the one-dimensional system $dy/dt = \frac{V_{max}u^r}{k_m + u^r} - kp$ and the memoryless system $u = \lambda y$, respectively.

The basic theorem for positive feedback analyzes an interconnection of two systems

$$\frac{dx_1}{dt} = f_1(x_1, u_1), \quad y_1 = h_1(x_1) \quad (2)$$

$$\frac{dx_2}{dt} = f_2(x_2, u_2), \quad y_2 = h_2(x_2) \quad (3)$$

which have increasing characteristics, denoted by “ k ” and by “ g ” respectively. (A special case is that in one of the systems is memoryless, for example if there are no state variables x_1 and y_1 is simply a static function $y_1(t) = k(u_1(t))$.)

For expository reasons (see Enciso and Sontag (2005b) for a generalization to high-dimensional inputs and outputs), we assume as in Angeli and Sontag (2004a) that the inputs and outputs of both systems are scalar: $m_1 = m_2 = p_1 = p_2 = 1$.

The “positive feedback interconnection” of the systems (2) and (3) is formally defined by letting the output of each of them serve as the input of the other ($u_2 = y_1 = “y”$ and $u_1 = y_2 = “u”$), as depicted in Fig. 15a. Let us now consider Fig. 15b, where we have plotted together k and the inverse of g . It is quite obvious that there is a bijective correspondence between the steady states of the feedback system and the intersection points of the two graphs. Moreover, let us attach labels to the intersection points between the two graphs as follows: a label “ S ” is placed at those points at which the slope of k is smaller than the slope of g^{-1} , and a label “ U ” if the slope of k is larger than the slope of g^{-1} . Note that in any interval between any consecutive two intersection points labeled S and U , the graph of g^{-1} is over the graph of k , and otherwise the graph of g^{-1} is under the graph of k . (We

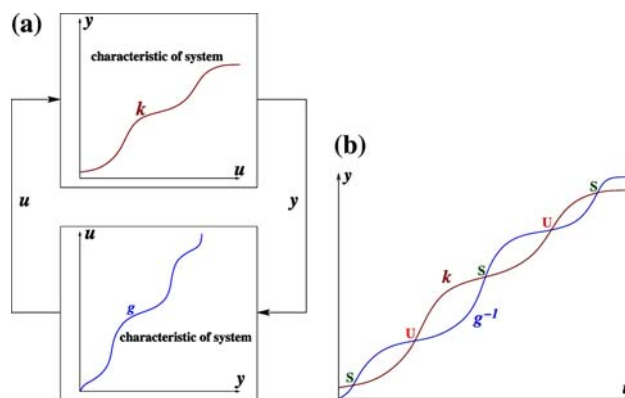


Fig. 15 (a) Positive feedback (b) characteristics

assume that the graphs don’t intersect tangentially.) By analogy with the previously considered simple example, one would expect that the points labeled S should correspond to stable states of the closed-loop system, while points labeled U should correspond to unstable states of the closed-loop system.

Indeed, let us consider the system $du/dt = y - g^{-1}(u)$, which has characteristic $u = g(y)$ when u is considered as an output and y as an input, connected in feedback with the system $y = k(u)$, seen as a memoryless system with u as input and y as output. The closed-loop system is:

$$\frac{du}{dt} = k(u) - g^{-1}(u)$$

and therefore $du/dt < 0$ in the intervals where the graph of g^{-1} is over the graph of k , which means that $u(t)$ will converge to a point labeled S when in an interval of the type “ (S, U) .” Conversely $du/dt > 0$ in the intervals where the graph of g^{-1} is under the graph of k , which means that $u(t)$ will also converge to a point labeled S when in an interval of the type “ (U, S) .” In summary, solutions move away from values of u corresponding to intersections labeled U and toward those corresponding to intersections labeled S . Similarly, $y(t) = k(u(t))$ converges to the points y associated to S ’s.

Of course, the systems may be more complicated than $du/dt = y - g^{-1}(u)$ and $y = k(u)$, so that the above paragraph does not constitute a proof. Nonetheless, a theorem to be explained below provides conditions insuring the validity of this argument. Before explaining the generalization, however, we provide a cautionary note, with the purpose of showing that intuition may sometimes fail.

A cautionary counterexample

We consider the following two-dimensional system (Angeli et al. 2004a):

$$\begin{aligned}\dot{x} &= x(-x + y) \\ \dot{y} &= 3y\left(-x + c + \frac{by^4}{k+y^4}\right)\end{aligned}$$

evolving on the first orthant $x > 0, y > 0$ (from now on, we use “ \dot{x} ” to denote time derivative), which provides a simplified model of the rate of change of the concentration of a protein x which may be degraded when in dimeric form (x^2 term) and whose formation is promoted by another protein having concentration y . In turn, the second protein is degraded by the first (term xy in second equation), and cooperative autocatalysis drives synthesis (last term in second equation). This is an activator/inhibitor or predator-prey system. We view this system as the unitary feedback system that results from setting $u = g(y) = y$ in the following open-loop system with input u and output y :

$$\begin{aligned}\dot{x} &= x(-x + y) \\ \dot{y} &= 3y\left(-x + c + \frac{bu^4}{k+u^4}\right).\end{aligned}$$

It is easy to verify that this open-loop system has the following characteristic:

$$k(u) = c + \frac{bu^4}{k+u^4}.$$

Figure 16 shows the plot of $y = k(u)$ (sigmoidal curve) together with the plot of $y = g^{-1}(u) = u$. The above discussion would then suggest that the points labeled I and III should correspond to stable states, and the point labeled II to an unstable state, with most trajectories converging to one of the two stable states. However, the phase plane of the closed loop system, as shown in Fig. 17, contains two unstable spiral points, in heteroclinic connections with a saddle, as well as a limit cycle. Thus, the conclusions fail.

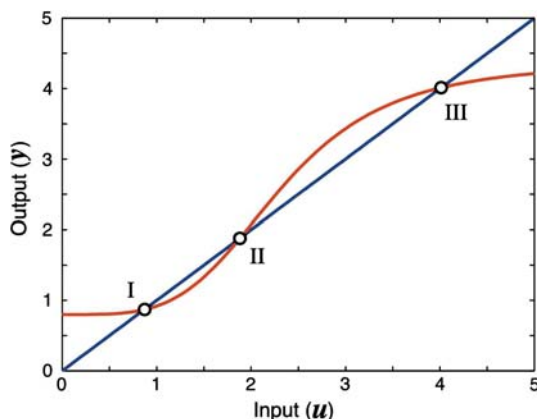


Fig. 16 Characteristics for counterexample

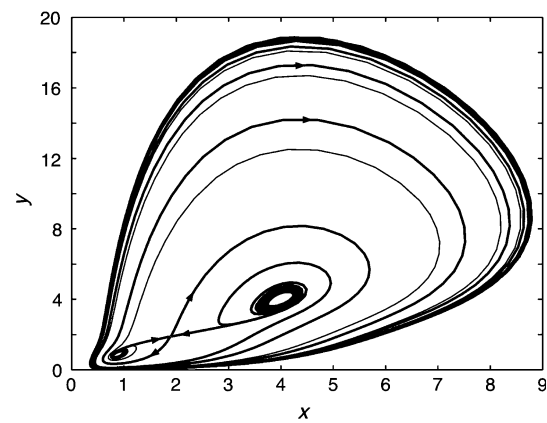


Fig. 17 Phase plane for counterexample

A general theorem

The above counterexample shows that the general interconnection theorem is not generally true. However, under the assumption that each open-loop system is monotone, together with reasonably mild technical conditions of transversality and “controllability” and “observability,” (the recent papers (Enciso and Sontag 2005a; G.A. Enciso and E.D. Sontag, in preparation) show that even these mild conditions can be largely dispensed with), the intuitive one-dimensional picture does generalize correctly. Suppose that we attached labels “ S ” and “ U ” as discussed earlier. Then, one can conclude that “almost all” (in a measure-theoretic sense or in a Baire-category sense) bounded solutions of the feedback system must converge to one of the steady-states corresponding to intersection points labeled with an S (Angeli and Sontag 2004a). The proof reduces ultimately to an application of Hirsch’s generic convergence theorem to the closed-loop system (the technical conditions insure strong monotonicity). However, the value-added is in the fact that stable states can be identified merely from the *one-dimensional* plot shown in Fig. 15b. (If each subsystem would have dimension just one, one can also interpret the result in terms of a simple nullcline analysis; see the Supplementary Section of Angeli et al. (2004a).) Of course, the system in the counterexample is not monotone; note the negative cycle in its influence graph Fig. 18.

We remark that the theorems remain true even if arbitrary delays are allowed in the feedback loop and/or if space-dependent models are considered and diffusion is allowed (see Sontag (2005) for a discussion). A new approach (Angeli 2006), based not on monotone theory but on a notion of “counterclockwise dynamics,” extends in a different direction the range of applicability of this methodology.

We wish to emphasize the potential practical relevance of this result (and others such as Angeli (2006)). The equations describing each of the systems are often poorly,

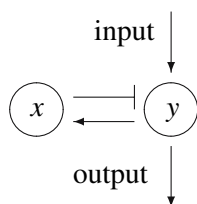


Fig. 18 Influence graph for counterexample

or not at all, known. But, as long as we can assume that each subsystem is monotone and monostable, we can use the information from the planar plots in Fig. 15b to understand the global dynamics of the closed-loop system, no matter how large the number of state variables. It is often said that the field of molecular systems biology is characterized by a *data-rich/data-poor* paradox: while on the one hand a huge amount of *qualitative* network (schematic modeling) knowledge is available for signaling, metabolic, and gene regulatory networks, on the other hand little of this knowledge is *quantitative*, at least at the level of precision demanded by most mathematical tools of analysis. On the other hand, input/output steady-state data (from a signal such as a ligand, to a reporter variable such as the expression of a gene monitored by GFP, or the activity of a protein measured by a Western blot) is frequently available. The problem of exploiting qualitative knowledge, and effectively integrating relatively sparse quantitative data, is among the most challenging issues confronting systems biology. The MIOS approach provides one way to combine these two types of data, hence addressing the “data-rich/data-poor” issue (Sontag 2004, 2005). When applicable, MIOS analysis allows one to combine the numerical information provided by the shape of the graphs of characteristics with the qualitative information given by (signed) network topology in order to predict global bifurcation behavior. This information is often easier to obtain from experimental data, at least in interpolated form, than kinetic constants (of which there may be a very large number). An analysis based on characteristics, when it can be done, is “robust” with respect to uncertainty in internal parameters of the system, and serves as a “qualitative-quantitative approach” to systems biology (Sontag 2005). In addition, characteristics are also a very powerful tool for

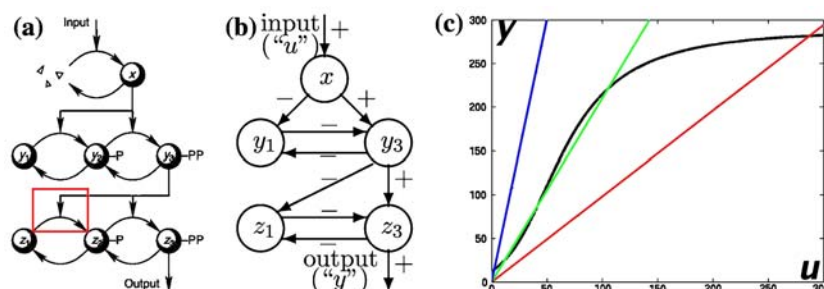
the purely mathematical analysis of existing models. Monotone systems with well-defined characteristics constitute a very well-behaved set of building blocks for arbitrary systems, as illustrated by the fact that cascades of such systems inherit the same properties (monotone, monostable response) and by the feedback theorems reviewed here, originally presented in the works (Angeli and Sontag 2004a; Angeli and Sontag 2003).

More discussion through an example: MAPK cascades

Mitogen-Activated Protein Kinase (MAPK) cascades are a ubiquitous “signaling module” in eukaryotes, involved in proliferation, differentiation, development, movement, apoptosis, and other processes (Huang and Ferrell Jr. 1996; Asthagiri and Lauffenburger 2001; Widmann et al. 1999). There are several such cascades, sharing the property of being composed of a cascade of three kinases. The basic rule is that two proteins, called generically MAPK and MAPKK (the last K is for “kinase of MAPK,” which is itself a kinase), are active when doubly phosphorylated, and MAPKK phosphorylates MAPK when active. Similarly, a kinase of MAPKK, MAPKKK, is active when phosphorylated. A phosphatase, which acts constitutively (that is, by default it is always active) reverses the phosphorylation. The biological model from Angeli et al. (2004a) and Huang and Ferrell (1996) is in Fig. 19b, where we wrote $z_i(t)$, $i = 1, 2, 3$ for MAPK, MAPK-P, and MAPK-PP concentrations and similarly for the other variables. The input represents an external signal to this subsystem (typically, the concentration of a kinase driving forward the reaction).

We make here the simplest assumptions about the dynamics, amounting basically to a quasi-steady-state approximation of enzyme kinetics. (For related results using more realistic, mass-action, models, see Angeli and Sontag (2007) and Angeli et al. (2006, 2007).) For example, take the reaction shown in the square in Fig. 19a. As y_3 (MAPKK-PP) facilitates the conversion of z_1 into z_2 (MAPK to MAPK-P), the rate of change dz_2/dt should include a term $\alpha(z_1, y_3)$ (and dz_1/dt has a term $-\alpha(z_1, y_3)$) for some (otherwise unknown) function α such that $\alpha(0, y_3) = 0$ and $\frac{\partial \alpha}{\partial z_1} > 0$, $\frac{\partial \alpha}{\partial y_3} > 0$ when $z_1 > 0$. (Nothing

Fig. 19 (a) MAPK cascades, (b) graph, (c) characteristic



happens if there is no substrate, but more enzyme or more substrate results in a faster reaction.) There will also be a term $+\beta(z_2)$ to reflect the phosphatase action. Similarly for the other species. The system as given would be represented by a set of seven ordinary differential equations (or reaction–diffusion PDE's, if spatial localization is of interest, or delay-differential equations, if appropriate).

This system is not monotone (at least with respect to any orthant cone), as is easy to verify graphically. However, as with many other examples of biochemical networks, the system is “monotone in disguise”, so to speak, in the sense that a judicious change of variables allows one to apply MIOS tools. (Far more subtle forms of this argument are key to applications to signaling cascades. A substantial research effort, not reviewed here because of lack of space, addresses the search for graph-theoretic conditions that allow one to find such “monotone systems in disguise”; see Sontag (2004, 2005) and Angeli et al. (2006) for references).

In this example, which in fact was the one whose study initially led to the definition of MIOS, the following conservation laws: $y_1(t) + y_2(t) + y_3(t) \equiv y_{\text{tot}}$ (total MAPKK) and $z_1(t) + z_2(t) + z_3(t) \equiv z_{\text{tot}}$ (total MAPK) hold true, assuming no protein turn-over. This assumption is standard in most of the literature, because transcription and degradation occur at time scales much slower than signaling. (There is very recent experimental data that suggests that turn-over might be fast for some yeast MAPK species. Adding turn-over would lead to a different mathematical model.) These conservation laws allow us to eliminate variables. The right trick is to eliminate y_2 and z_2 . Once we do this, and write $y_2 = y_{\text{tot}} - y_1 - y_3$ and $z_2 = z_{\text{tot}} - z_1 - z_3$, we are left with the variables x, y_1, y_3, z_1, z_3 . For instance, the equations for z_1, z_3 look like:

$$\begin{aligned}\frac{dz_1}{dt} &= -\alpha(z_1, y_3) + \beta(z_{\text{tot}} - z_1 - z_3) \\ \frac{dz_3}{dt} &= \gamma(z_{\text{tot}} - z_1 - z_3, y_3) - \delta(z_3)\end{aligned}$$

for appropriate increasing functions $\alpha, \beta, \gamma, \delta$. The equations for the remaining variables are similar. The graph, ignoring, as usual, self-loops (diagonal of Jacobian), is shown in Fig. 19b. This graph has no negative undirected loops, showing that the (reduced) system is monotone. A consistent spin assignment (including the top input node and the bottom output node) is shown in Fig. 20. It is also true that this system has a well-defined monostable state space response (characteristic); there is no space to discuss the proof here, so we refer the reader to the original papers (Angeli and Sontag 2003, 2004b).

Positive and negative feedback loops around MAPK cascades have been a topic of interest in the biological

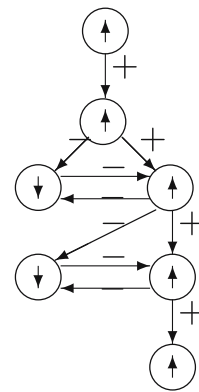


Fig. 20 Consistent assignment for simple MAPK cascade model

literature. For example, see Ferrell and Machleder (1998) and Bhalla et al. (2002) for positive feedback and Kholodenko (2000) and Shvartsman et al. (2000) for negative feedback. Since we know that the system is monotone and has a characteristic, MIOS theory as described here can indeed be applied to the example. We study next the effect of a positive feedback $u = g y$ obtained by “feeding back” into the input a scalar multiple g of the output. (This is a somewhat unrealistic model of feedback, since feedbacks act for example by enhancing the activity of a kinase. We pick it merely for illustration of the techniques.)

The theorem does not require actual equations for its applicability. All that is needed is the knowledge that we have a MIOS, and a plot of its characteristic (which, in practice, would be obtained from interpolated experimental data). In order to illustrate the conclusions, on the other hand, it is worth discussing a particular set of equations. We take equations and parameters from Angeli et al. (2004a), Sontag (2004, 2005):

$$\begin{aligned}\frac{dx}{dt} &= -\frac{v_2 x}{k_2 + x} + v_0 u + v_1 \\ \frac{dy_1}{dt} &= \frac{v_6 (Y - y_1 - y_3)}{k_6 + (Y - y_1 - y_3)} - \frac{k_3 x y_1}{k_3 + y_1} \\ \frac{dy_3}{dt} &= \frac{k_4 x (Y_{\text{tot}} - y_1 - y_3)}{k_4 + (Y - y_1 - y_3)} - \frac{v_5 y_3}{k_5 + y_3} \\ \frac{dz_1}{dt} &= \frac{v_{10} (z_{\text{tot}} - z_1 - z_3)}{k_{10} + (z - z_1 - z_3)} - \frac{k_7 y_3 z_1}{k_7 + z_1} \\ \frac{dz_3}{dt} &= \frac{v_8 y_3 (z_{\text{tot}} - z_1 - z_3)}{k_8 + (z - z_1 - z_3)} - \frac{v_9 z_3}{k_9 + z_3}\end{aligned}$$

with output z_3 . Specifically, we will use the following parameters: $v_0 = 0.0015$, $v_1 = 0.09$, $v_2 = 1.2$, $v_3 = 0.064$, $v_4 = 0.064$, $v_5 = 5$, $v_6 = 5$, $v_7 = 0.06$, $v_8 = 0.06$, $v_9 = 5$, $v_{10} = 5$, $y_{\text{tot}} = 1,200$, $z_{\text{tot}} = 300$, $k_2 = 200$, $k_3 = 1,200$, $k_4 = 1,200$, $k_5 = 1,200$, $k_6 = 1,200$, $k_7 = 300$, $k_8 = 300$, $k_9 = 300$, $k_{10} = 300$. (The units are: totals in nM (mol/cm³), v 's in nM s⁻¹ and s⁻¹, and k 's in nM.)

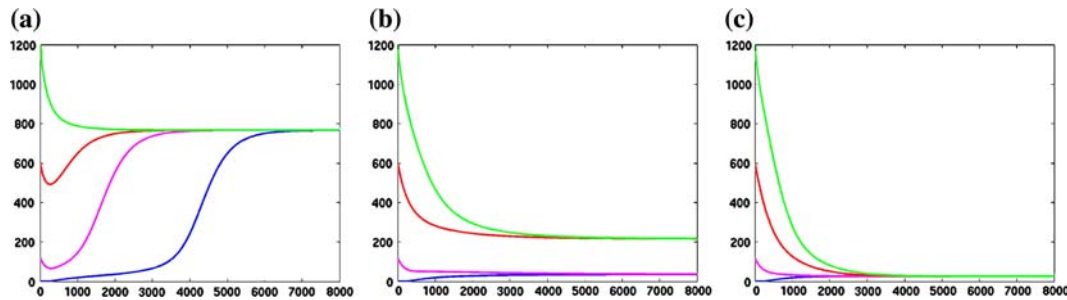


Fig. 21 (a),(b),(c) y_3 , $g = 1/0.98, 1/2.1, 1/6$

With these choices, the steady-state step response is the sigmoidal curve shown in Fig. 19c, where y is the output z_3 . We plotted in the same figure the inverse g^{-1} of the characteristic of the feedback system, in this case just the linear mapping $y = (1/g)u$, for three typical “feedback gains” ($g = 1/0.98, 1/2.1, 1/6$).

For $g = 1/0.98$ (line of slope 0.98 when plotting y against u), there should be a unique stable state, with a high value of the output $y = z_3$, and trajectories should generically converge to it. Similarly, for $g = 1/2.1$ (line of slope 2.1) there should be two stable states, one with high and one with low $y = z_3$, with trajectories generically converging to one of these two, because the line intersects at three points, corresponding to two stable and one unstable state (exactly as in the discussion concerning the simple protein formation/degradation sigmoidal example in Fig. 13). Finally, for $g = 1/6$ (line of slope 6), only the low- y stable state should persist. Fig. 21a–c shows plots of the hidden variable $y_3(t)$ (MAPKK-PP) for several initial states, confirming the predictions. The same convergence results are predicted if there are delays in the feedback loop, or if concentrations depend on location in a convex spatial domain. Results for reaction–diffusion PDE’s and delay-differential systems are discussed in Sontag (2005), and simulation results for this example are also provided there.

We may plot the steady-state value of y , under the feedback $u = gy$, as the gain g is varied, Fig. 22a.

This resulting complete bifurcation diagram showing points of saddle-node bifurcation can be also completely determined just from the characteristic, with no need to know the equations of the system. Relaxation oscillations may be expected under such circumstances if a second, slower, feedback loop is used to negatively adapt the gain as a function of the output. Reasons of space preclude describing a very general theorem, which shows that indeed, relaxation oscillations can be guaranteed in this fashion: see Gedeon and Sontag (2007) for technical details, and Sontag (2005) for a more informal discussion. Fig. 22b shows a simulation confirming the theoretical prediction (details in Sontag (2005) and Gedeon and Sontag (2007)).

Negative feedback and possible oscillations

A different set of results apply to inhibitory or negative feedback interconnections of two MIOS systems (2)–(3). A convenient mathematical way to define “negative feedback” in the context of monotone systems is to say that the orders on inputs and outputs are inverted (example: an inhibition term of the form $\frac{V}{K+y}$ as usual in biochemistry). Equivalently, we may incorporate the inhibition into the output of the second system (3), which is then seen as an *anti-monotone* I/O system, and this is how we proceed from now on. See Fig. 23a. We emphasize that the closed-

Fig. 22 (a) Bifurcation diagram and relaxation (b) oscillation (y_3)

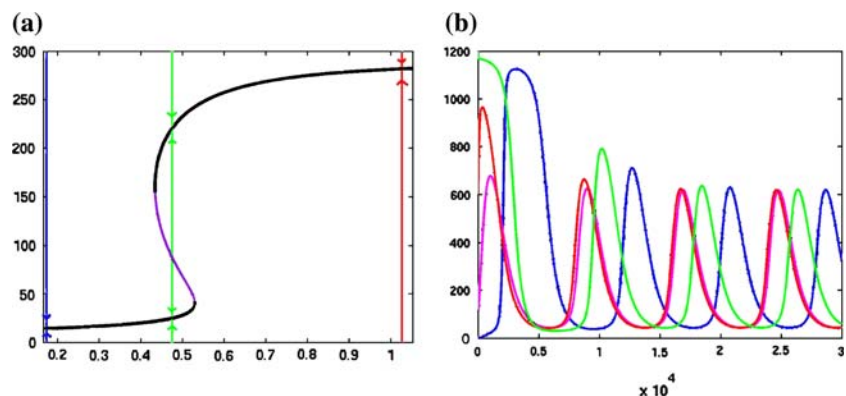
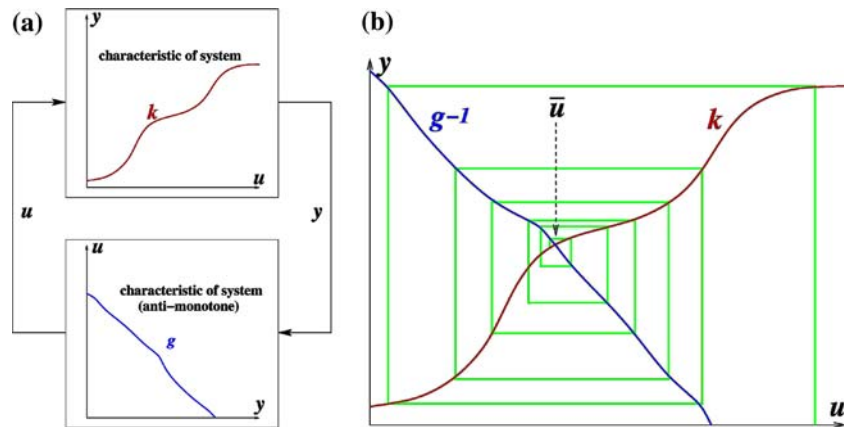


Fig. 23 (a) Negative feedback and (b) characteristics



loop systems that result are *not* monotone, at least with respect to any known order.

The original theorem, from Angeli and Sontag (2003), is as follows. We assume once more that inputs and outputs are scalar ($m = p = 1$; see Enciso and Sontag (2006) for generalizations). We once again plot together k and g^{-1} , as shown in Fig. 23b. Consider the following discrete iteration:

$$u_{i+1} = (g \circ k)(u_i).$$

Then, if solutions of the closed-loop system are bounded and if this iteration has a globally attractive fixed point \bar{u} , as shown in Fig. 23b, then the feedback system has a globally attracting steady-state. (An equivalent condition, see Enciso and Sontag (2006), is that the iteration have no nontrivial period-two orbits.) We call this result a *small gain theorem* (“SGT”), because of its analogy to concepts in control theory.

It is easy to see that *arbitrary delays* may be allowed in the feedback loop. In other words, the feedback could be of the form $u(t) = y(t - h)$, and such delays (even with $h = h(t)$ time varying or even state-dependent, as long as $t - h(t) \rightarrow \infty$ as $t \rightarrow \infty$) do not destroy global stability of the closed loop. In Enciso et al. (2006), we have now shown also that *diffusion* does not destroy global stability either. In other words, a reaction–diffusion system (Neumann boundary conditions) whose reaction can be modeled in the shown feedback configuration, has the property that all solutions converge to a (unique) uniform in space solution. This is not immediately obvious, since standard parabolic comparison theorems do not immediately apply to the feedback system, which is not monotone.

Example: MAPK cascade with negative feedback

As with the positive feedback theorem, an important feature is applicability to highly uncertain systems. As long as the component systems are known to be MIOS, the knowledge of I/O response curves and a planar analysis are

sufficient to conclude GAS of the entire system, which may have an arbitrarily high dimension. For example, suppose we take a feedback like $u = a + b/(c + z_3)$, with a graph as shown in Fig. 24a, which also shows the characteristic and a convergent discrete 1-d iteration (Sontag 2005). Then, we are guaranteed that all solutions of the closed-loop system converge to a unique steady-state, as confirmed by the simulations in Fig. 24b, which shows the concentrations of the active forms of the kinases.

Example: testosterone model

This example is intended to show that even for a classical mathematical biology model, a very simple application of the result in Angeli and Sontag (2003) gives an interesting conclusion. The concentration of testosterone in the blood of a healthy human male is known to oscillate periodically with a period of a few hours, in response to similar oscillations in the concentrations of the luteinising hormone (LH) secreted by the pituitary gland, and the luteinising hormone releasing hormone (LHRH), normally secreted by the hypothalamus (see Cartwright and Husain 1986). The well-known textbook (Murray 2002) (and its previous editions) presents this process as an example of a biological oscillator, and proposes a model to describe it, introducing

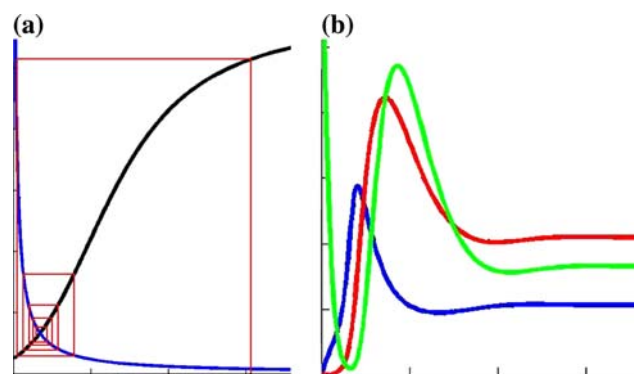


Fig. 24 Inhibition: (a) spiderweb and (b) simulation

delays in order to obtain oscillations. (Since the textbook was written, the physiological mechanism has been much further elucidated, and this simple model is now known not to be correct. However, we want merely to illustrate a point about mathematical analysis.) The equations are:

$$\begin{aligned}\dot{R} &= \frac{A}{K+T} - b_1 R \\ \dot{L} &= g_1 R - b_2 L \\ \dot{T} &= g_2 L(t-\tau) - b_3 T\end{aligned}$$

(R, L, T = concentrations of hormones luteinising hormone releasing, luteinising, and testosterone, τ = delay). The system may be seen as the feedback connection of the MIOS system

$$\begin{aligned}\dot{R} &= u - b_1 R \\ \dot{L} &= g_1 R - b_2 L \\ \dot{T} &= g_2 L - b_3 T\end{aligned}$$

with the inhibitory feedback $u(t) = g(T - \tau) = A/(K + T(t - \tau))$ after moving the delay to the loop (without loss of generality). The characteristic is linear, $T = k(u) = \frac{g_1 g_2}{b_1 b_2 b_3} u$, so $g \circ k$ is a fractional transformation $S(u) = \frac{p}{q+u}$. Since such a transformation has no period-two cycles, global stability follows. (For arbitrary, even time-varying, delays.) This contradicts the existence of oscillations claimed in Murray (2002) for large enough delays. (See Enciso and Sontag (2004), which also explains the error in Murray (2002).)

Example: Lac operon

The study of *E. Coli* lactose metabolism has been a topic of research in mathematical biology since Jacob and Monod's classical work which led to their 1955 Nobel Prize. For this example, we look at the subsystem modeled in Mahaffy and Savev (1999). The lac operon induces production of permease and β -gal, permease makes the cell membrane more permeable to lactose, and genes are activated if lactose present; lactose is digested by the enzyme β -gal, and the other species are degraded at fixed rates. (In this model from Mahaffy and Savev (1999), lactose and isolactose are identified, and catabolic repression by glucose via cAMP is ignored.) Delays arise from translation of permease and β -gal. The equations are:

$$\begin{aligned}\dot{x}_1(t) &= g(x_4(t-\tau)) - b_1 x_1(t) \quad \text{lac operon mRNA} \\ \dot{x}_2(t) &= x_1(t) - b_2 x_2(t) \quad \beta\text{-galactoside permease} \\ \dot{x}_3(t) &= r x_1(t) - b_3 x_3(t) \quad \beta\text{-galactosidase} \\ \dot{x}_4(t) &= S x_2(t) - x_3(t) x_4(t) \quad \text{lactose}\end{aligned}$$

with $g(x) := (1 + Kx^\rho)/(1 + x^\rho)$, $K > 1$, and the Hill exponent ρ representing a cooperativity effect. (All delays have been lumped into one.) We view this system as a negative feedback loop, where $u = x_1$, $v = x_4$, of a MIOS system (details in Enciso and Sontag (2006)). Since there are two inputs and outputs, now we must study the two-dimensional iteration

$$(u, v) \mapsto (g \circ k)(u, v) = \left[\frac{g(v)}{b}, \frac{S b_1 b_3 u}{r b_2 g(v)} \right].$$

Based on results on rational difference equations from Kulenovic and Ladas (2002), one concludes that there are no nontrivial 2-periodic orbits, provided that $\rho < (\sqrt{K} + 1)/(K - 1)$, for arbitrary b_1, b_2, b_3, r, S . Hence, by the theorem, there is a unique steady-state of the original system, which is GAS, even when arbitrary delays are present.

These and other conditions are analyzed in Enciso and Sontag (2006), where it is also shown that the results from Mahaffy and Savev (1999) are recovered as a special case. Among other advantages of this approach, besides generalizing the result and giving a conceptually simple proof, we have (because of Enciso et al. (2006)) the additional conclusion that also for the corresponding reaction-diffusion system, in which localization is taken account of, the same globally stable behavior can be guaranteed.

Example: Circadian oscillator

As a final example of the negative feedback theorem, we pick Goldbeter's (1995, 1996) original model of the molecular mechanism underlying circadian rhythms in *Drosophila*. (In this oversimplified model, only *per* protein is considered; other players such as *tim* are ignored.) PER protein is synthesized at a rate proportional to its mRNA concentration. Two phosphorylation sites are available, and constitutive phosphorylation and dephosphorylation occur with saturation dynamics, at maximum rate v_i 's and with Michaelis constants K_i . Doubly phosphorylated PER is degraded, also satisfying saturation dynamics (with parameters v_d, k_d), and it is translocated to the nucleus with rate constant k_1 . Nuclear PER inhibits transcription of the *per* gene, with a Hill-type reaction of cooperativity degree n and threshold constant K_I , and mRNA is produced, and translocated to the cytoplasm, at a rate determined by a constant v_s . Additionally, there is saturated degradation of mRNA (constants v_m and k_m). The model is (P_i = *per* phosphorylated at i sites, P_N = nuclear *per*, M = *per* mRNA):

$$\begin{aligned}
\dot{M} &= v_s \frac{K_I}{K_I + P_N^n} - v_m \frac{M}{k_m + M} \\
\dot{P}_0 &= k_s M - V_1 \frac{P_0}{K_1 + P_0} + V_2 \frac{P_1}{K_2 + P_1} \\
\dot{P}_1 &= V_1 \frac{P_0}{K_1 + P_0} - V_2 \frac{P_1}{K_2 + P_1} - V_3 \frac{P_1}{K_3 + P_1} + V_4 \frac{P_2}{K_4 + P_2} \\
\dot{P}_2 &= V_3 \frac{P_1}{K_3 + P_1} - V_4 \frac{P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{k_d + P_2} \\
\dot{P}_N &= k_1 P_2 - k_2 P_N.
\end{aligned}$$

Parameters are chosen exactly as in Goldbeter's original paper, except that the rate v_s of mRNA translocation to the cytoplasm is taken as a bifurcation parameter. The value $v_s = 0.76$ from Goldbeter (1995) gives oscillatory behavior. On the other hand, we may break up the system into the M and P_i, P_N subsystems. Each of these can be shown to be MIOS and have a characteristic. (The existence of a characteristic for the P -subsystem is nontrivial, and involves the application of Smillie's Theorem (Smillie 1984) for strongly monotone tridiagonal systems, and more precisely, repeated application of a proof technique in Smillie (1984) involving "eventually monotonicity" of state variables.) When $v_s = 0.4$, the discrete iteration is graphically seen to be convergent (see Fig. 25a), so the theorem guarantees global asymptotic stability even when arbitrary delays are introduced in the feedback. Bifurcation analysis on delay length and v_s indicates that local stability will fail for somewhat larger values. Using again the graphical test, we observe that for $v_s = 0.5$ there appears limit cycle for the *discrete* iteration on characteristics, see Fig. 25b. This suggests that oscillations may exist in the full nonlinear differential equation, at least for appropriate delays lengths. Indeed, the simulation in Fig. 25c displays such oscillations (Angeli and Sontag 2004b, c), and a Hopf bifurcation can be shown to exist (Angeli and Sontag 2007).

A counterexample

We now provide a (non-monotone) system as well as a feedback law $u = g(y)$ so that: the system has a well-defined and increasing characteristic k , and the discrete

iteration $u^+ = g(k(u))$ converges globally, and solutions of the closed-loop system are bounded, yet a stable limit-cycle oscillation exists in the closed-loop system. This establishes, by means of a simple counterexample, that *monotonicity* of the open-loop system is an essential assumption in the MIOS negative feedback theorem. Thus, robustness of the conclusion of stability is only guaranteed with respect to uncertainty that preserves monotonicity of the system. Using language from control theory, the idea underlying the construction is very simple. The open-loop system is linear, and has the following transfer function:

$$W(s) = \frac{-s + 1}{s^2 + (0.25)s + 1}.$$

Since the DC gain of this system is $W(0) = 1$, and the system is stable, there is a well-defined and increasing characteristic $k(u) = u$. However, a negative feedback gain of $1/2$ destabilizes the system, even though the discrete iteration $u^+ = (-1/2)u$ is globally convergent. (The H_∞ gain of the system is, of course, larger than 1, and therefore the standard small-gain theorem does not apply.) In state-space terms, we use this system:

$$\begin{aligned}
\dot{x}_1 &= (-1/4)x_1 - x_2 + 2u \\
\dot{x}_2 &= x_1 \\
y &= (1/2)(x_2 - x_1).
\end{aligned}$$

Note that, for each constant input $u \equiv u_0$, the solution of the system converges to $(0, u_0/2)$, and therefore the output converges to u_0 , so indeed the characteristic k is the identity. We only need to modify the feedback law in order to make solutions of the closed-loop globally bounded. For the feedback law we pick $g(x) = -0.5\text{sat}(y)$, where $\text{sat}(\cdot) := \text{sign}(\cdot)\min\{1, |\cdot|\}$ is a saturation function. The only equilibrium of the closed-loop system is at $(0,0)$.

The discrete iteration is

$$u^+ = -(1/2)\text{sat}(u).$$

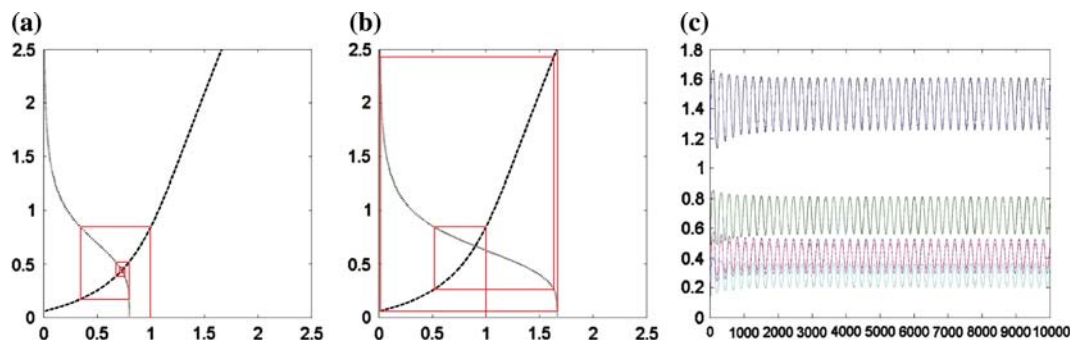


Fig. 25 (a) Convergent iteration, (b) divergent iteration, (c) oscillations

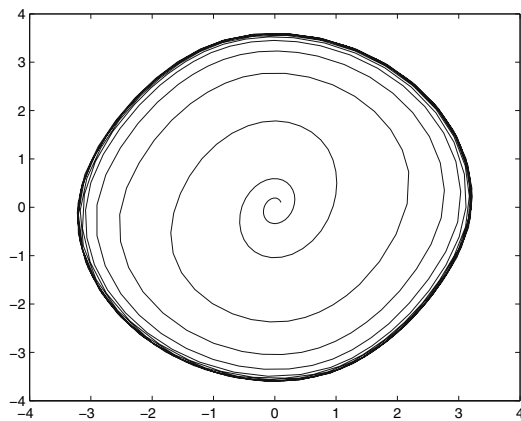


Fig. 26 Limit cycle in counterexample

With an arbitrary initial condition u_0 , we have that $u_1 = -(1/2)\text{sat}(u_0)$, so that $|u_1| \leq 1/2$. Thus $u_k = (-1/2)u_{k-1}$ for all $k \geq 2$, and indeed $u_k \rightarrow 0$ so global convergence of the iteration holds.

However, global convergence to equilibrium fails for the closed-loop system, and in fact there is a periodic solution. Indeed, note that trajectories of the closed loop system are bounded, because they can be viewed as solutions of a stable linear system forced by a bounded input. Moreover, since the equilibrium is a repelling point, it follows by the Poincaré-Bendixson Theorem that a periodic orbit exists. Fig. 26 is a simulation showing a limit cycle.

Conclusions

There is a clear need in systems biology to study robust structures and to develop robust analysis tools. The theory of monotone systems provides one such tool. Interesting and nontrivial conclusions can be drawn from (signed) network structure alone, which is associated to purely stoichiometric information about the system, and ignores fluxes.

Associating a graph to a given system, we may define spin assignments and consistency, a notion that may be interpreted also as non-frustration of Ising spin-glass models. Every species in a monotone system (one whose graph is consistent) responds with a consistent sign to perturbations at every other species. This property would appear to be desirable in biological networks, and, indeed, there is some evidence suggesting the near-monotonicity of some natural networks. Moreover, “near”-monotone systems might be “practically” monotone, in the sense of being monotone under disjoint environmental conditions.

Dynamical behavior of monotone systems is ordered and “non-chaotic.” Systems close to monotone may be

decomposed into a small number of monotone subsystems, and such decompositions may be usefully employed to study non-monotone dynamics as well as to help detect bifurcations even in monotone systems, based only upon sparsenumerical data, resulting in a sometimes useful model-reduction approach.

Acknowledgments Much of the author’s work on I/O monotone systems was done in collaboration with David Angeli, as well as Patrick de Leenheer, German Enciso, Bhas kar Dasgupta, and Hal Smith. The author also wishes to thank Moe Hirsch, Reka Albert, Tom Knight, Avi Maayan, Alex van Oudenaarden, and many others, for useful comments and suggestions regarding the material discussed here.

References

- Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the *drosophila* segment polarity genes. *J Theoret Biol* 223:1–18
- Allwright DJ (1977) A global stability criterion for simple control loops. *J Math Biol* 4:363–373
- Anderson I (2002) Combinatorics of finite sets. Dover Publications, Mineola, NY
- Angeli D (2006) Systems with counterclockwise input–output dynamics. *IEEE Trans Automatic Control* 51:1130–1143
- Angeli D, Sontag ED (2003) Monotone control systems. *IEEE Trans Automat Control* 48(10):1684–1698. Errata are here: [http://www.math.rutgers.edu/~\(tilde\)sontag/FTPDIR/angeli-sontag-monotone-TAC03-typos.txt](http://www.math.rutgers.edu/~(tilde)sontag/FTPDIR/angeli-sontag-monotone-TAC03-typos.txt)
- Angeli D, Sontag ED (2004a) Multi-stability in monotone input/output systems. *Syst Control Lett* 51(3–4):185–202
- Angeli D, Sontag ED (2004b) Interconnections of monotone systems with steady-state characteristics. In: *Optimal control, stabilization and nonsmooth analysis*, vol 301 of *Lecture Notes in Control and Information Science*. Springer, Berlin, pp 135–154
- Angeli D, Sontag ED (2004c) An analysis of a circadian model using the small-gain approach to monotone systems. In: *Proc. IEEE Conf. Decision and Control*, Paradise Island, Bahamas, Dec. IEEE Publications, pp 575–578
- Angeli D, Sontag ED (2007) Analysis of a circadian model using the small-gain approach to monotone systems. In: *IEEE transactions automation and control*, Special Issue on Systems Biology, January 2007 (submitted). Preprint version in *arXiv q-bio.QM/0701018*, 14 Jan 2007
- Angeli D, Sontag ED (2007) Translation-invariant monotone systems, and a global convergence result for enzymatic futile cycles. *Nonlinear Anal Ser B: Real World Appl* (to appear)
- Angeli D, Ferrell JE, Sontag ED (2004a) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Natl Acad Sci USA* 101(7):1822–1827, February 2004. A revision of Suppl. Fig.7(b) is here:[http://www.math.rutgers.edu/~\(tilde\)sontag/FTPDIR/nullclines-f-g-REV.jpg](http://www.math.rutgers.edu/~(tilde)sontag/FTPDIR/nullclines-f-g-REV.jpg); and typos can be found here: [http://www.math.rutgers.edu/~\(tilde\)sontag/FTPDIR/angeli-ferrell-sontag-pnas04-errata.txt](http://www.math.rutgers.edu/~(tilde)sontag/FTPDIR/angeli-ferrell-sontag-pnas04-errata.txt).
- Angeli D, de Leenheer P, Sontag ED (2004b) A small-gain theorem for almost global convergence of monotone systems. *Syst Control Lett* 52(5):407–414
- Angeli D, de Leenheer P, Sontag ED (2006) On the structural monotonicity of chemical reaction networks. In: *Proc. IEEE*

- Conf. Decision and Control, San Diego, Dec. 2006, p WeA01.2. IEEE, 2006
- Angeli D, de Leenheer P, Sontag ED (2007) A Petri net approach to the study of persistence in chemical reaction networks. *Math Biosci* (to appear). Also arXiv q-bio.MN/068019v2, 10 Aug 2006
- Aracena J, Demongeot J, Goles E (2004) On limit cycles of monotone functions with symmetric connection graph. *Theoret Comput Sci* 322(2):237–244
- Asthagiri AR, Lauffenburger DA (2001) A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol Prog* 17:227–239
- Bagowski CP, Ferrell JE Jr (2001) Bistability in the JNK cascade. *Curr Biol* 11:1176–1182
- Bagowski CP, Besser J, Frey CR, Ferrell JE Jr (2003) The JNK cascade as a biochemical switch in mammalian cells: ultrasensitive and all-or-none responses. *Curr Biol* 13:315–320
- Barahona F (1982) On the computational complexity of Ising spin glass models. *J Phys A Math Gen* 15:3241–3253
- Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J* 20:2528–2535
- Bhalla US, Ram PT, Iyengar R (2002) Map kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* 297:1018–1023
- Cartwright M, Husain MA (1986) A model for the control of testosterone secretion. *J Theoret Biol* 123:239–250
- Chaves M, Sontag ED, Dinerstein RJ (2004) Steady-states of receptor-ligand dynamics: a theoretical framework. *J Theoret Biol* 227(3):413–428
- Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *J Theoret Biol* 235(3):431–449
- Cinquin O, Demongeot J (2002) Positive and negative feedback: striking a balance between necessary antagonists. *J Theoret Biol* 216:229–241
- Clarke BL (1980) Stability of complex reaction networks. In: Prigogine I, Rice SA (eds) *Advances in chemical physics*. John Wiley, New York, pp 1–215
- Clive Maxfield (2006) How to invert three signals with only two not gates (and *no* xor gates). Technical report, <http://www.mobilehandsetdesignline.com>
- Costanzo MC, Crawford ME, Hirschman JE, Kranz JE, Olsen P, Robertson LS, Skrzypek MS, Braun BR, Hopkins KL, Kondu P, Lengieza C, Lew-Smith JE, Tillberg M, Garrels JI (2001) YPDTM, PombePDTM and WormPDTM: model organism volumes of the BioKnowledgeTM Library, an integrated resource for protein information. *Nucl Acids Res* 29(1):75–79
- Craciun G, Feinberg M (2005) Multiple equilibria in complex chemical reaction networks: I. The injectivity property. *SIAM J Appl Mathematics* 65:1526–1546
- Craciun G, Feinberg M (2006) Multiple equilibria in complex chemical reaction networks: II. the species-reactions graph. *SIAM J Appl Math* 66:1321–1338
- Cross FR, Archambault V, Miller M, Klovstad M (2002) Testing a mathematical model of the yeast cell cycle. *Mol Biol Cell* 13:52–70
- Dancer EN (1998) Some remarks on a boundedness assumption for monotone dynamical systems. *Proc AMS* 126:801–807
- DasGupta B, Enciso GA, Sontag ED, Zhang Y (2007) Algorithmic and complexity aspects of decompositions of biological networks into monotone subsystems. *BioSystems* (to appear)
- De Leenheer P, Malisoff M (2006) A small-gain theorem for monotone systems with multivalued input-state characteristics. *IEEE Trans Automat Control* 51:287–292
- De Simone C, Diehl M, Junger M, Mutzel P, Reinelt G, Rinaldi G (1995) Exact ground states of Ising spin glasses: new experimental results with a branch and cut algorithm. *J Stat Phys* 80:487–496
- DeAngelis DL, Post WM, Travis CC (1986) *Positive feedback in natural systems*. Springer-Verlag, New York
- Devaney R (1989) *An introduction to chaotic dynamical systems*, 2nd edn. Addison-Wesley, Redwood City
- Doyle JC, Francis B, Tannenbaum A (1990) *Feedback control theory*. MacMillan Publishing Co.
- Edelstein-Keshet L (2005) *Mathematical models in biology*. SIAM, Philadelphia
- Enciso GA, Sontag ED (2004) On the stability of a model of testosterone dynamics. *J Math Biol* 49(6):627–634
- Enciso GA, Sontag ED (2005a) A remark on multistability for monotone systems ii. In: *Proc. IEEE conf. decision and control*, Seville, Dec. 2005, IEEE Publications, pp 2957–2962
- Enciso GA, Sontag ED (2005) Monotone systems under positive feedback: multistability and a reduction theorem. *Syst Control Lett* 54(2):159–168
- Enciso GA, Sontag ED (2006) Global attractivity, I/O monotone small-gain theorems, and biological delay systems. *Discrete Contin Dyn Syst* 14(3):549–578
- Enciso GA, Smith HL, Sontag ED (2006) Non-monotone systems decomposable into monotone systems with negative feedback. *J Differ Eqs* 224:205–227
- Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors – i. the deficiency zero and deficiency one theorems. *Chem Eng Sci* 42:2229–2268
- Feinberg M (1991) Some recent results in chemical reaction network theory. In: Aris R, Aronson DG, Swinney HL (eds) *Patterns and dynamics in reactive media*, IMA, Vol Math Appl 37. Springer, Berlin, p 4370
- Feinberg M (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch Rational Mech Anal* 132:311–370
- Feinberg M, Horn FJM (1974) Dynamics of open chemical systems and algebraic structure of underlying reaction network. *Chem Eng Sci* 29:775–787
- Fenichel N (1979) Geometric singular perturbation theory for ordinary differential equations. *J Differ Eqs* 31:53–98
- Ferrell JE Jr (1996) Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem Sci* 21:460–466
- Ferrell JE Jr, Machleder EM (1998) The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science* 280:895–898
- Ferrell JE Jr, Xiong W (2001) Bistability in cell signaling: how to make continuous processes discontinuous, and reversible processes irreversible. *Chaos* 11:227–236
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *escherichia coli*. *Nature* 403:339–342
- Gedeon T (1998) Cyclic feedback systems. *Mem Am Math Soc* 134:1–73
- Gedeon T, Sontag ED (2007) Oscillations in multi-stable monotone systems with slowly varying feedback. *J Differ Eqs* (to appear)
- Gilbert EN (1954) Lattice theoretic properties of frontal switching functions. *J Math Phys* 33:57–67
- Goemans M, Williamson D (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J ACM* 42:1115–1145
- Goldbeter A (1995) A model for circadian oscillations in the drosophila period protein (per). *Proc Roy Soc Lond B* 261:319–324
- Goldbeter A (1996) *Biochemical oscillations and cellular rhythms*. Cambridge University Press, Cambridge

- Gouze J-L (1988) A criterion of global convergence to equilibrium for differential systems. application to Lotka-Volterra systems. Technical Report RR-0894, INRIA
- Gouze JL (1998) Positive and negative circuits in dynamical systems. *J Biol Syst* 6:11–15
- Gouze J-L, Haderl KP (1994) Order intervals and monotone flow. *Nonlinear World* 1:23–34
- Haderl K, Glas D (1983) Quasimonotone systems and convergence to equilibrium in a population genetics model. *J Math Anal Appl* 95:297–303
- Hale JK (1988) Asymptotic behavior of dissipative systems. Amer Math Soc, Providence
- Harary F (1953) On the notion of balance of a signed graph. *Michigan Math J* 2:143–146
- Hastings S, Tyson J, Webster D (1977) Existence of periodic solutions for negative feedback cellular control systems. *J Differ Eqs* 25:39–64
- Hess P, Poláčik P (1993) Boundedness of prime periods of stable cycles and convergence to fixed points in discrete monotone dynamical systems. *SIAM J Math Anal* 24:1312–1330
- Hirsch M (1983) Differential equations and convergence almost everywhere in strongly monotone flows. *Contemp Math* 17:267–285
- Hirsch MW (1984) The dynamical systems approach to differential equations. *Bull AMS* 11:1–64
- Hirsch M (1985) Systems of differential equations that are competitive or cooperative ii: convergence almost everywhere. *SIAM J Math Anal* 16:423–439
- Hirsch M (1989) Convergent activation dynamics in continuous-time networks. *Neural Networks* 2:331–349
- Hirsch M, Smith HL (2005) Monotone dynamical systems. In: *Handbook of differential equations, ordinary differential equations* (second volume). Elsevier, Amsterdam
- Horn FJM (1974) The dynamics of open reaction systems. In: *Mathematical aspects of chemical and biochemical problems and quantum chemistry*. Proc. SIAM-AMS Sympos Appl Math New York, 1974, pp 125–137. Amer Math Soc, Providence, 1974. SIAM-AMS Proceedings, vol. VIII
- Horn FJM, Jackson R (1972) General mass action kinetics. *Arch Rational Mech Anal* 49:81–116
- Huang C-YF, Ferrell JE Jr (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci USA* 93:10078–10083
- Hüffner F, Betzler N, Niedermeier R (2007) Optimal edge deletions for signed graph balancing. In: *Proceedings of the 6th workshop on experimental algorithms (WEA07)*, June 6–8, 2007, Rome. Springer-Verlag
- Istrail S (2000) Statistical mechanics, three-dimensionality and NP-completeness: I. Universality of intractability of the partition functions of the Ising model across non-planar lattices. In: *Proceedings of the 32nd ACM symposium on the theory of computing (STOC00)*. ACM Press, pp 87–96
- Jiang JF (1994) On the global stability of cooperative systems. *Bull London Math Soc* 6:455–458
- Jones CKRT (1994) Geometric singular perturbation theory. In: *Dynamical systems (Montecatini Terme)*, Lect Notes in Math, vol 1696. Springer-Verlag, Berlin
- Kauffman SA (1969a) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theoret Biol* 22:437–467
- Kauffman SA (1969b) Homeostasis and differentiation in random genetic control networks. *Nature* 224:177–178
- Kauffman SA, Glass K (1973) The logical analysis of continuous, nonlinear biochemical control networks. *J Theoret Biol* 39:103–129
- Keener JP, Sneyd J (1998) *Mathematical physiology*. Springer-Verlag, New York
- Khalil HK (2002) *Nonlinear systems*, 3rd edn. Prentice Hall, Upper Saddle River, NJ
- Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* 267:1583–1588
- Kulenovic MRS, Ladas G (2002) *Dynamics of second order rational difference equations*. Chapman & Hall/CRC, New York
- Laurent M, Kellershohn N (1999) Multistability: a major means of differentiation and evolution in biological systems. *Trends Biochem Sci* 24:418–422
- de Leenheer P, Angeli D, Sontag ED (2005) On predator–prey systems and small-gain theorems. *Math Biosci Eng* 2(1):25–42
- de Leenheer P, Angeli D, Sontag ED (2007) Monotone chemical reaction networks. *J Math Chem* (to appear)
- Lewis J, Slack JM, Wolpert L (1977) Thresholds in development. *J Theoret Biol* 65:579–590
- Lisman JE (1985) A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc Natl Acad Sci USA* 82:3055–3057
- Mahaffy J, Savev ES (1999) Stability analysis for a mathematical model of the *lac* operon. *Quart Appl Math* LVII:37–53
- Mallet-Paret J, Smith HL (1990) The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *J Dyn Differ Eqs* 2:367–421
- Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* 110:11980–11985
- Mangan S, Zaslaver A, Alon U (2003) The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* 334:197–204
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298:824–827
- Minsky ML (1967) *Computation: finite and infinite machines*. Prentice-Hall, Englewood Cliffs, NJ
- Monod J, Jacob F (1961) Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb Symp Quant Biol* 26:389–401
- Murray JD (2002) *Mathematical biology*, I, II: an introduction. Springer-Verlag, New York
- Novic A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA* 43:553–566
- Othmer HG (1976) The qualitative dynamics of a class of biochemical control circuits. *J Math Biol* 3:53–78
- Plahte E, Mestl T, Omholt WS (1995) Feedback circuits, stability and multistationarity in dynamical systems. *J Biol Syst* 3:409–413
- Poláčik P, Tereščák I (1992) Convergence to cycles as a typical asymptotic behavior in smooth strongly monotone discrete-time dynamical systems. *Arch Rational Mech Anal* 116:339–360
- Poláčik P, Tereščák I (1993) Exponential separation and invariant bundles for maps in ordered banach spaces with applications to parabolic equations. *J Dyn Differ Eqs* 5:279–303
- Pomerening JR, Sontag ED, Ferrell JE (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of *cdc2*. *Nat Cell Biol* 5(4):346–351, April 2003. Supplementary materials 2–4 are here: <http://www.math.rutgers.edu/~tilde/sontag/FTPDIR/pomerening-sontag-ferrell-a-additional.pdf>
- Ptashne M (1992) *A Genetic switch: phage λ and higher organisms*. Cell Press and Blackwell Scientific Publications, Cambridge MA
- Rapp PE (1975) A theoretical investigation of a large class of biochemical oscillations. *Math Biosci* 25:165–188
- Reddy VN, Mavrovouniotis ML, Liebman MN (1993) Petri net representations in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol* 1:328–336
- Remy E, Mosse B, Chaouiya C, Thieffry D (2003) A description of dynamical graphs associated to elementary regulatory circuits. *Bioinformatics* 19(Suppl 2):ii172–ii178

- Schneider H, Vidyasagar M (1970) Cross-positive matrices. *SIAM J Numer Anal* 7:508–519
- Segel LA (1984) Modeling dynamic phenomena in molecular and cellular biology. Cambridge University Press, Cambridge
- Sepulchre R, Jankovic M, Kokotović PV (1997) Constructive nonlinear control. Springer-Verlag, London
- Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, Tyson JJ, Sible JC (2003) Hysteresis drives cell-cycle transitions in *xenopus laevis* egg extracts. *Proc Natl Acad Sci USA* 100:975–980
- Shvartsman SY, Wiley HS, Lauffenburger DA (2000) Autocrine loop as a module for bidirectional and context-dependent cell signaling. Technical report, MIT Chemical Engineering Department
- Smale S (1976) On the differential equations of species in competition. *J Math Biol* 3:5–7
- Smillie J (1984) Competitive and cooperative tridiagonal systems of differential equations. *SIAM J Math Anal* 15:530–534
- Smith HL (1987) Oscillations and multiple steady states in a cyclic gene model with repression. *J Math Biol* 25:169–190
- Smith HL (1991) Convergent and oscillatory activation dynamics for cascades of neural nets with nearest neighbor competitive or cooperative interactions. *Neural Networks* 4:41–46
- Smith H (1995) Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems, Mathematical Surveys and Monographs, vol 41, AMS, Providence, RI
- Snoussi EH (1998) Necessary conditions for multistationarity and stable periodicity. *J Biol Syst* 6:3–9
- Sontag ED (1998) Mathematical control theory. Deterministic finite-dimensional systems, vol 6 of texts in applied mathematics, 2nd edn. Springer-Verlag, New York
- Sontag ED (1999) Stability and stabilization: discontinuities and the effect of disturbances. In: Nonlinear analysis, differential equations and control, Montreal, QC, 1998, vol 528 of NATO Sci Ser C Math Phys Sci. Kluwer Acad Publ, Dordrecht, pp 551–598
- Sontag ED (2004) Some new directions in control theory inspired by systems biology. *IEEE Proc Syst Biol* 1:9–18
- Sontag ED (2005) Molecular systems biology and control. *Eur J Control* 11(4–5):396–435
- Tereščák I (1996) Dynamics of c^1 smooth strongly monotone discrete-time dynamical system. Technical report, Comenius University, Bratislava
- Thieme HR (1992) Convergence results and a Poincaré-Bendixson trichotomy for asymptotically autonomous differential equations. *J Math Biol* 30:755–763
- Thomas R (1981) On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer Ser Synerget* 9:180–193
- Thomas R, D’ari R (1990) Biological feedback. CRC Press, Boca Raton
- Thomas R, Kaufman M (2001) Multistationarity, the basis of cell differentiation and memory. i. structural conditions of multistationarity and other nontrivial behavior. *Chaos* 11:170–179
- Thron CD (1991) The secant condition for instability in biochemical feedback-control.1. The role of cooperativity and saturability. *Bull Math Biol* 53:383–401
- Tyson J, Othmer HG (1978) The dynamics of feedback control circuits in biochemical pathways. *Prog Theoret Biol* 5:1–60
- Tyson JJ, Chen K, Novak B (2003) Sniffers, buzzers, toggles, and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15:221–231
- Volkman P (1972) Gewöhnliche differentialungleichungen mit quasimonoton wachsenden funktionen in topologischen vektorräumen. *Math Z* 127:157–164
- Walcher S (2001) On cooperative systems with respect to arbitrary orderings. *J Math Anal Appl* 263:543–554
- Walter W (1970) Differential and Integral Inequalities. Springer-Verlag, Berlin
- Wang L, Sontag ED (2006a) Almost global convergence in singular perturbations of strongly monotone systems. In: Positive Systems. Springer-Verlag, Berlin, pp 415–422. Lecture Notes in Control and Information Sciences, vol 341, Proceedings of the second multidisciplinary international symposium on positive systems: theory and applications (POSTA 06) Grenoble, France
- Wang L, Sontag ED (2006b) A remark on singular perturbations of strongly monotone systems. In: Proc. IEEE conf. decision and control, San Diego, Dec. 2006, p WeB10.5. IEEE
- Widmann C, Spencer G, Jarpe MB, Johnson GL (1999) Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev* 79:143–180
- Zaslavsky T (1998) Bibliography of signed and gain graphs. *Electron J Combin* DS8
- Zevedei-Oancea I, Schuster S (2003) Topological analysis of metabolic networks based on petri net theory. *In Silico Biol* 3:0029